

The Effects of Computer Assisted Interactive Fiction
on the Development of Reading Comprehension
in Grade Five Students

Robert F. Whitelock, B.A., B.Ed.

Department of Graduate and Undergraduate Studies in
Education

(Submitted in partial fulfillment of
the requirements for the degree of
Master of Education)

College of Education
Brock University
St. Catharines, Ontario

© April, 1988

Acknowledgements

I would like to thank Dr. David Young, Assistant Superintendent, Fort McMurray School District 2833, for his support and encouragement, not only with this project, but throughout my graduate program. Special thanks go to Ms. Paula Hanson, TRC Consultant, whose assistance in conducting literature searches and interest in the project were both helpful and inspirational.

I would also like to thank the staff and students of Westview Public School in Fort McMurray, Alberta, who participated in the experiment. Without the co-operation of Mrs. Valerie MacIntyre, Mr. Terry Harrison, and their respective classes, this study could not have been undertaken. Ms. Leigh Anne Willard, Guidance Counsellor at Westview, was particularly helpful in providing background information and materials with regard to the testing involved.

I would like to express my deep appreciation to those who have worked with me through the process of preparing this document. Dr. A. Wheeler of Brock University encouraged the initial undertaking, and has provided much needed interest and support throughout the entire project. Dr. G. Samson's assistance in the processing of the data obtained from the experiment was critical, and his interest, encouragement and late night telephone calls were deeply appreciated.

Lastly, I would like to thank Mr. John Miller, who provided opportunities during working hours to travel to St. Catharines in conjunction with the research and defense of this thesis, my colleague Mr. Frank Sweet for his painstakingly thorough review of the draft text, and my wife, Mary, who proofed and initially critiqued the writing; whose encouragement, patience and caring throughout my graduate program have made this document possible.

The successes of this study are shared with all of the persons above, and many others unmentioned. Any errors or omissions belong to me, alone.

Abstract

This study was undertaken in order to determine the effects of playing computer based text adventure games on the reading comprehension gains of students. Forty-five grade five students from one elementary school were randomly assigned to experimental and control groups, and were tested with regard to ability, achievement and reading skills. An experimental treatment, consisting of playing computer based interactive fiction games of the student's choice for fifteen minutes each day over an eight-week period, was administered. A comparison treatment engaged the control group in sustained silent reading of materials of the student's choice for an equal period of time. Following the experimental period all students were post-tested with an alternate form of the pre-test in reading skills, and gain scores were analysed. It was found that there were no significant differences in the gain scores of the experimental and control groups for overall reading comprehension, but the experimental group showed greater gains than the control group in the structural analysis reading sub-skill. Extreme variance in the data made generalization very difficult, but the findings indicated a potential for computer based interactive fiction as a useful tool for developing reading skills. The great need for further research in the same vein was highlighted in the conclusions.

TABLE OF CONTENTS

Acknowledgements	ii
Abstract	iv
List of Tables	vii
List of Illustrations	ix
 Chapter 1 -- Introduction	 1
Statement of the Problem	2
Rationale	2
Background	6
Assumptions and Limitations	7
Definition of Terms	8
Reading comprehension	8
Computer based learning	9
Computer based interactive fiction	10
Hypotheses	10
Summary	11
 Chapter 2 -- Review of the Literature	 13
Reading Comprehension	13
Schema theory	14
Story grammars	16
Student interest	17
Instructional effects	19
Silent reading	20
Uninterrupted sustained silent reading	23
Computers and Reading	25
Pedagogy	27
Computer delivery of text	28
Research on CAI and achievement	29
Research on CAI and reading achievement	31
Summary of computers in reading instruction	32
Computer Based Interactive Fiction	34
Summary	37
Endnotes	40
 Chapter 3 -- Experimental Design	 41
Design	42
Background	42
Sample	43
Method	43
Treatment	44
Instrumentation	46
Descriptive pre-testing	46
Reading comprehension testing	47
Other measures	48
Summary	49
Endnotes	49

Chapter 4 -- Analysis of the Data	51
Introduction	51
Sample Characteristics	51
Ability	52
Achievement	53
Others	53
Experimental Conditions	54
Findings of the Hypotheses	55
Hypothesis 1	55
Hypothesis 2	56
Hypothesis 3	57
Hypothesis 4	58
Discussion of the Findings	59
Observational data	59
Variables	64
Age	64
Sex	65
Ability	66
Vocabulary achievement	67
Reading achievement	68
Spelling achievement	69
Language achievement	70
Text difficulty	72
Attendance	74
Homeroom class	75
Summary of comparisons by variables	77
Summary	80
Endnotes	81
Chapter 5 -- Summary and Conclusions	82
Summary	82
Main Features of the Method	83
Main Findings of the Study	84
Conclusions Based on the Findings	85
Quality of the data	85
Validity of the findings	89
Generalizability	90
Potential Design Improvements	91
Research Arising from the Study	95
Conclusion	97
Endnotes	98
References	99
Uncited Statistical References	105
Instruments Used in the Experiment	105
Interactive Fiction Used in the Experiment	105
Appendix A	106
Appendix B	107

LIST OF TABLES

Table 1	Comparison of Population Data between Experimental and Control Groups	52
Table 2	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills between Experimental and Control Groups	55
Table 3	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Ability Groups	58
Table 4	Summary of Hypotheses and Decisions	60
Table 5	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Age Groups	65
Table 6	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Sex.....	66
Table 7	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Vocabulary Achievement Groups....	67
Table 8	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Reading Achievement Groups.....	69
Table 9	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Spelling Groups	70
Table 10	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Language Skills Groups.....	71
Table 11	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Text Difficulty Groups.....	72
Table 12	Correlation Co-efficients between Text Difficulty and Selected Variables	74
Table 13	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Attendance Groups.....	75
Table 14	Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Homeroom Class.....	76
Table 15	Correlation Co-efficients for Selected Variables by Reading Components.....	77
Table 16	T-test Results Comparing Mean Scores of Experimental and Control Groups in Inferential Comprehension by Selected Variables	79

Table 17 Summary of Hypotheses and Decisions	84
Table 18 A Comparison of the Size of Standard Deviations to Means of Gain Scores	84

LIST OF ILLUSTRATIONS

Figure 1 Frequency Distributions of Gain Scores in Total Comprehension for Experimental and Control Groups	56
Figure 2 Comparison of Reading Comprehension and Reading Sub-skills by Variable Characteristics Yielding Significantly Higher Gain Scores	79

Chapter 1 -- Introduction

Since the late 1970's the number of microcomputers in schools in North America has been increasing rapidly. Over the same period there has not, however, been a corresponding increase in our understanding of the best use of computers in classrooms. Purchases of microcomputers have been based more on societal beliefs about the instructional power of computers and parental concerns about the future employability of children than any pedagogical understanding of the role that microcomputers should play in the education of students (Becker, 1984; Bear, 1984; Winkler, Shavelson, Stasz, Robyn, & Feibel, 1985).

Despite increasing availability, recent surveys of computer use in classrooms have revealed that a very small percentage of the learning of any given student is accomplished through computer related instruction (Bork, 1984). Computer based learning most frequently involves computer assisted instruction (CAI), specifically drill and practice exercises (Oakes and Schneider, 1984). These exercises are usually little different from the paper and pencil variety (Bork, 1984).

Computer based interactive fiction, or adventure games, represent a departure from regular computer related activities. The student role-plays a character within a story

which unfolds through the student's own actions, keeping the reader closely involved with the narrative at all times. This has led to speculation that playing adventure games may have a significant positive impact on student reading skills and interest level (Kuechle, 1983; Newman, 1984; Jarchow and Montgomery, 1985; Little, 1985; Grabe & Grabe, 1985; Brady, 1986; Anderson-Inman, 1987).

Statement of the Problem

One of the methods of developing student reading comprehension in school involves giving students opportunities to read. Many teachers provide students with time to read silently each day in order to foster an interest in reading and provide reading practice. Computer based interactive fiction involves reading, and a host of other mental and physical activities. What would be the effects on students' reading comprehension of substituting equal time playing computer based interactive fiction for silent reading of paper-based reading materials? The purpose of this study was to examine the impact of student participation in computer text adventure games on the reading comprehension of grade five students over an eight week period.

Rationale

Historically, most new forms of educational technology have been considered to be best suited to individualization, behavioral objectives, task and concept analysis and the rigorous use of reinforcement schedules (Parry, Thorkildsen,

Blery, & Macfarlane, (1985 - 86). A number of surveys of schools in North America have determined that the predominant form of computer based learning is drill-and-practice and tutorial CAI (Oakes and Schneider, 1984; Becker, 1984) which fits this characterization very closely. Kulik, Bangert and Williams (1983) reviewed studies undertaken with students in grades 6 - 12 and concluded that CAI is effective in raising student achievement, shortening time taken to learn and improving student attitudes towards learning. These findings have been echoed in both secondary school studies (Samson, Niemiec, Weinstein and Walberg, 1985) and elementary level research (Niemiec, Samson, Weinstein and Walberg, 1987). They are not supported by Jameson, Suppes and Wells (1974) and Clark (1983), however, who reviewed studies of instructional media and concluded that no one particular form of instructional technology could be considered superior to any other, and that the activities of the teacher involved were the key to student achievement. It can therefore be concluded that at present the true effects of computer based learning are not clearly understood.

There are currently two models related to the development of reading skills in students, the reductionist and whole language approaches. The reductionist model divides reading into a collection of sub-skills. Instruction involves practicing specifically identified skills related to reading through objective-specific drills. Computer managed drill-and-practice exercises fit naturally into this model of reading

instruction, providing individualized practice of discrete objectives in reading sub-skills (Newman, 1984).

Virtually all computer assisted reading programs developed since the 1960's have been of the drill-and-practice or tutorial kinds, centring around reading strands such as letter identification, sight word vocabulary, spelling patterns, phonics and simple comprehension. The majority of computer based reading programs emphasize Skinnerian reinforcement schedules in a manner similar to programmed instruction (Bradley, 1985).

The alternative to the reductionist model sees reading instruction as an integrated part of language learning. Rather than work on individual skills, the language experience (or whole language) approach engages students in reading activities for the purpose of actively searching for meaning (Grabe & Grabe, 1985). The aim of this approach is to treat reading as an integrated part of all language arts, ensuring that the student can easily make the link between printed words and personal experience that is required for true comprehension to take place (Crafton, 1982).

CAI tutorial and drill-and-practice routines do not lend themselves easily to the whole language approach, as their inherent reliance on discrete sub-skill learning is not one of the intentions of the approach. A number of authors have suggested that an alternative use of computers in reading instruction that does support the methods of the whole language approach may be having students play adventure

games (Kuechle, 1983; Newman, 1984; Jarchow and Montgomery, 1985; Little, 1985; Grabe & Grabe, 1985; Brady, 1986; Anderson-Inman, 1987).

Adventure games are a computerized version of interactive fiction, a unique genre of literature in which the reader is required to make choices and direct the actions of at least one character in the story (Grabe and Grabe, 1985). Paper novels of this kind have been available for a number of years. Since the advent of the microcomputer, many old and new interactive stories have been turned into computer programs.

Computer based interactive fiction is readily available commercially. These programs exist for all makes of computers, and at a wide variety of levels of playing difficulty. The stories inherent in the adventures range from adaptations of well-known novels and children's tales to newly devised science fiction, historical or romance-based stories (Grabe & Grabe, 1985). Participation in an adventure game involves assuming the role of at least one of the characters in the game, reading information as it unfolds on the monitor screen, making decisions about what the character should do, and typing in words and phrases which bring about the furtherance of the plot. This sequence is repeated hundreds to thousands of times from the start of an adventure game to its ending.

The need for further research in all areas related to computer based learning is stated by virtually all

researchers. There is very little research verifying the cost-effective use of computers in schools (Bear, 1984). According to Ragsdale (1985), this is consistent with the development of all new forms of communications technology, in that we only come to a true understanding of the nature, best uses and effects of the technology once it has been in existence for some time. Bradley (1985) suggests that there is a specific need to identify through research which reading skills are best learned through computer intervention.

Bass, Ries and Sharpe (1986) raise a concern with respect to the generalizability of research conducted in the early years of CAI due to the highly clinical nature of the studies conducted at the time. Early studies were conducted on university campuses, using software created for the study running on mainframe computers, and supervised by researchers knowledgeable in both the study and the equipment. Since the early 1980's schools have been using microcomputers and commercially available software. The context under which the study was done may thus be a source of confounding. As a result, research undertaken prior to the introduction of the microcomputer may not be applicable to current situations.

Background

Prior to this study, a review of research related to computer based instruction and student achievement was undertaken (Whitelock, 1987). While that review was not

specifically related to the subject of the current study, many of its conclusions are related to it. First, the number of studies found which were related to computer based learning and student achievement of all kinds was small. This would indicate that the effects of computer based learning on achievement are not well understood in any particular context, including reading. Second, an overwhelming number of the studies located were related to CAI drill and practice or tutorial exercises only. No studies of non-conventional computer based learning, such as using interactive fiction, were found. The conclusion issued a call for such studies. This study is a move in the recommended research direction.

There was a distinct difference found in the results of studies conducted prior to and subsequent to the introduction of microcomputers in schools, supporting the contention of Bass, Ries and Sharpe (1986). This casts doubt on the usefulness of pre-1980 studies in reviewing the literature of the field for purposes of understanding a study based on microcomputers.

Assumptions and Limitations

In designing this study it was assumed that (i) engaging in daily silent reading would produce measurable gains over pre-test scores in reading comprehension, and that (ii) measurable gains would be made within an eight-week period of study.

In keeping with the concerns of Bass, Ries and Sharpe (1986), and as a result of the findings of the review described above, this study was designed to involve children working individually and concurrently with microcomputers and commercially available software. In addition, the assessment instruments were selected from common commercially-available standardized tests, such as the Stanford Diagnostic Reading Test, in order to facilitate replication of the study.

This study was undertaken within a single school, using two actual classes of students. No screening or selection of the sample occurred prior to commencement of the study. As a result, any findings will be generalizable only to the extent to which the sample is representative of the population.

Definition of Terms

Reading comprehension. Reading comprehension involves finding meaning in written text (Grabe and Grabe, 1985). Geoffrion and Geoffrion (1984/85) state that growth in reading comprehension involves growth in four elements: vocabulary (breadth of words known plus recognition of multiple meanings of words), language skills (ability to infer meaning), schemata (applying prior knowledge) and manipulative skills (integrating the above appropriately in reading a given text). Smith (1982) describes reading as the process of asking questions of a text, and comprehension as the degree to which the questions asked are answered. The

extent to which the questions are answered depends on the reader's prior knowledge of the matter dealt with by the text, ability to decode the words of the text itself and understanding of the grammatical structures in which the words are rendered. This is similar to the view held by Geoffrion and Geoffrion. As a result, reading comprehension is defined for the purposes of this study as a measure of the student's ability to understand text through decoding of the printed material and relating it to prior experience.

Computer based learning. A wide variety of expressions exist in the literature with respect to learning which is in some way accomplished with or through a computer. Computer assisted instruction (CAI) involves the provision of direct instruction to students by a computer, either in the form of drill-and-practice exercises, tutorials, or simulations (Oakes and Schneider, 1984). CAI implies that all instruction takes place at the computer carrel, and any other media are interfaced with the computer there (McCombs, Eschenbrenner and O'Neil, 1973). CAI can also involve the use of instructional games, interactive reading and some aspects of using the computer as a tool, insofar as they assist in the learning of traditional tasks. The most prevalent forms of CAI are drill-and-practice and tutorials, as they most closely resemble traditional instructional methods (Oakes and Schneider, 1984).

Computer managed learning (CML) involves the use of the computer to track and maintain the learning or a

student or a group of students. The learning takes place largely away from the computer, and the computer is used in order to assess, monitor and keep records. When CAI and CML are both used simultaneously, the activity is appropriately called computer based learning (CBL) (McCombs et al., 1973), or alternately computer based instruction (CBI).

Computer based interactive fiction. Computer based interactive fiction refers to text only and mixed graphics-and-text stories which are presented on a computer and which require the reader to actively participate in the story (Layton, 1987). The participant role-plays at least one character in the story. It involves reading text and making decisions which drive the plot of the story to some conclusion (Jarchow and Montgomery, 1985).

Hypotheses

This study involves a direct substitution of computer reading for the reading of paper-based materials. As a result, it is hypothesized that:

- 1) students engaged in computer based interactive fiction will not achieve significantly different gain scores in overall reading comprehension than students who engaged in uninterrupted sustained silent reading for equivalent time periods;
- 2) students engaged in computer based interactive fiction will not achieve significantly different gain scores in any reading comprehension sub-skill than students who engaged

in uninterrupted sustained silent reading for equivalent time periods;

3) students of different ability levels engaged in computer based interactive fiction will not vary significantly in overall reading comprehension gains as measured by pre- and post-test reading comprehension scores;

4) students of different ability levels engaged in computer based interactive fiction will not vary significantly in gains in any reading comprehension sub-skill as measured by pre-and post-test reading comprehension sub-skill scores.

Summary

The purpose of this study was to determine the effects of engaging in computer based interactive fiction on students' reading comprehension relative to the reading of paper based books. It arises out of the general need for further research in computer based learning, and specifically from the need for studies related to learning using computers in non-traditional CAI methods.

This paper contains four chapters beyond this introduction. Chapter 2 will review the literature which is related to the study and provide the theoretical base for its undertaking. Chapter 3 will describe the experimental design, sample and instrumentation used in the study. Chapter 4 will present the findings of the research and discuss the hypotheses of the study. Chapter 5 will discuss the nature of the findings, conclusions which can be drawn from the

findings, and suggest research which arises from this study. Following these chapters will be a bibliography of all sources used in the paper, and an appendix containing samples of the text read by students engaged in text adventures.

Chapter 2 -- Review of the Literature

Reading Comprehension

Reading comprehension has been defined as a measure of the student's ability to understand text through decoding of the printed material and relating it to prior experience. Decoding is the mechanical process of letter and word recognition. There is no clear agreement on how readers actually go about decoding text. Nicholson (1986) described the theoretical debate that has ensued between the two existing perspectives. On one hand are those who believe that reading is a top-down process, in which the reader uses the context of the sentence to understand the meanings of words which are unknown, relying as little as possible on decoding skills. The alternate, or bottom-up view, is that readers decode all words from memory, and only use context when all decoding options have failed. Research conducted by Goodman and Gough (cited in Nicholson), the proponents of the top-down and bottom-up theories respectively, has failed to provide conclusive evidence in favour of either theory.

Regardless of how a reader goes about decoding the text of a document, comprehension is achieved when the reader understands what the decoded words mean in the way that the author has put them together. Cognitive psychology suggests that this occurs when the reader is able to identify elements of the text being read with personal experiences or

knowledge systems. The personal background and experiences of the reader are thus crucial to the comprehension of the material being read. This is the role of schemata in reading (Crafton, 1982).

Schema Theory. Schemata are the building blocks of understanding (Rumelhart, 1980). A schema is an information structure in the memory which represents generic concepts¹. According to schema theory, the reader constantly evaluates what is being read against schemata in the memory, seeking those which fit the information being decoded as closely as possible. When the reader is able to link the decoded message with a schema, the material being read is comprehended.

Rumelhart suggests that if readers lack the appropriate schemata, they are likely to have problems remembering what they have read. It is also possible for readers having the appropriate schemata to have comprehension problems, if the author's cues are not powerful enough to trigger the needed schema.

Freebody and Anderson (1983) conducted two studies to determine the effects of text cohesion and schema availability on the comprehension of social studies passages that varied in vocabulary difficulty. The subjects of both studies were sixth grade students. The first study utilized two deliberately modified forms of one social studies text which enhanced and eroded the cohesion of the text. No significant main effects for either vocabulary difficulty or text cohesion with respect

to comprehension were found. The second study involved familiar and unfamiliar versions of two themes, in which vocabulary and cohesion were essentially constant. It was found that the availability of schemata had a very significant positive effect on student comprehension of the passages used in the study.

The impact of clues in text and prior knowledge of subject matter on student comprehension of printed texts was examined by Mazor and Yussen (1983). The subjects were students in grades 2 through 4. The first experiment, which did not control for student prior knowledge, found that helpful clues in the text resulted in improved student performance on passage-specific questions.

The second experiment controlled student prior knowledge by giving the experimental group information relevant to the test story prior to the test story being read. It was found that prior knowledge had a more significant effect than irrelevant prior knowledge, helpful clues and unhelpful clues in passage-specific test scores.

There thus appears to be a strong relationship between student prior knowledge and comprehension of written passages (Crafton, 1982; Rumelhart, 1980). The work of Mazor and Yussen (1983) indicated, however, that teachers can provide the necessary prior knowledge to students before embarking on a text which might fall outside the experience of students. This is consistent with the opinion of Crafton (1982).

Story grammars. One piece of prior knowledge which may have an impact on student comprehension of text is story grammar. Story grammar refers to the structure of any narrative form of communication, oral or written (Smith, 1982). All stories contain a setting (time and space), a plot (including a problem, goals and a method of resolution), and a resolution (Schmitt and O'Brien, 1986). Students who are aware of the grammar of a particular story are better able to comprehend the story than those who are not (Geoffrion and Geoffrion, 1984/85). Smith (1982) states that story grammars are so important that readers who do not recognize the grammar or 'genre' of a story being read will not only have problems comprehending it, but will, when asked to recall elements of the story, phrase them within a story grammar that they are comfortable with rather than try to piece together the story's own. Story grammars are, therefore, schemata which may be available to the reader (Schmitt and O'Brien, 1986).

Schmitt and O'Brien (1986) reviewed research on story grammars and reading comprehension in discussing the recent trend towards the direct teaching of story grammars as a means of promoting comprehension. They found that most studies of story grammars utilize cognitive measures that are based upon the concept of story grammars, making their validity questionable. They also found that the models of story grammars recommended for teaching are not generalizable enough to be of any real use outside the type of

story that they are designed for. As a result, Schmitt and O'Brien recommended that story grammars be adapted by teachers to fit the stories being taught in class, rather than being taught discretely as an end in themselves.

Student interest. Personal interest is a very powerful motivator for learning. Schemata are created by the individual in response to interesting phenomena in the surroundings. Interesting phenomena are usually events or circumstances which are similar enough to existing schemata to be identifiable, but not so similar as to be understood at first glance (Good and Brophy, 1977). It can thus be expected that a student's comprehension of reading material considered interesting would be greater than that for material considered uninteresting.

Research on the relationship between interest in reading materials and comprehension was reviewed by Asher (1980). He reported that there have been few such studies, and those that exist may have been biased, either because interest has been tested for after the reading had taken place, or the study was undertaken by making gender or age based assumptions about what text would be interesting rather than seeking student opinion. As a result, Asher (and others) undertook three separate studies which eliminated these biases.

Asher found that student desire to read materials of interest and comprehension of materials of interest was significantly higher than for materials which were not

considered interesting. Two of the three studies reported higher comprehension results for boys than for girls due to high-interest material. Girls were found to perform better than boys on low-interest material, but not on high-interest material. A noteworthy additional finding was that boys tended to score less well than girls on reading achievement tests. Asher postulated that this may have been a result of low-interest material on the tests, rather than any real achievement difference.

The findings of Asher (1980) were confirmed in more recent studies undertaken by Deutsch (1985) and Guzzetti (1984). Deutsch examined interest in reading material, learning and comprehension using modified versions of a fable, testing for interest rating, comprehension rating and measured reading comprehension. Significant positive main effects were found for interest rating and comprehension rating, and interest rating and measured reading comprehension.

Guzzetti (1984) studied the relationship between comprehension and content areas through an examination of miscue correction. It was found that students were able to correct a significantly larger number of miscues in materials in which the student indicated an interest than other materials.

Research has thus shown that the availability of schemata, knowledge of story grammars and personal interest are positive predictors of reading comprehension.

The factors examined thus far explain, in cognitive terms, the processes by which readers come to comprehend text, but they do not provide much insight in terms of areas in which classroom practices can be adapted to improve the reading comprehension of students.

Instructional effects. Leinhardt, Zigmond and Cooley (1981) conducted a study to determine the significant instructional factors that influence the development of reading skills in elementary aged children. The sample consisted of 105 children ranging from 6 to 12 years of age who were diagnosed as learning disabled, in eleven separate classrooms in the Pittsburgh area. The IQ range of the sample was 63 to 153. It included a 2:1 ratio of boys to girls and whites to non-whites.

The method involved pre- and post-testing using the Spacke Diagnostic Reading Test and the Wide Range Achievement Test Reading Level I, and post-testing with the Comprehensive Test of Basic Skills reading subtest. In addition, the teachers and students involved were observed regularly by trained observers. Seventeen different variables were identified which might affect reading.

The study found that gains in post-test scores were significantly influenced by silent reading time, overlap of class instruction with the test contents, and the pre-test result. Variables found not significant were oral reading and indirect reading (reading done in passing). Gains in total reading behaviours were found to be influenced by teacher

instruction, reinforcement, the pre-test and cognitive press, in descending order of significance. In concluding, Leinhardt et al. recommended that reducing the amount of nonproductive time students spend in school and increasing the amount of silent reading time by as little as 5 to 10 minutes per day will result in significant improvement in student reading ability.

The study by Leinhardt et al. was conducted using subjects who were classified as learning disabled, but the conclusions and recommendations are not so made. Further research would be necessary to determine if these findings were generalizable to the population at large.

Silent reading. The issue of the role of silent reading in the development of reading comprehension bears further discussion. Historically, silent reading has been considered to reflect the true reading process, as silent reading is done for the purpose of obtaining meaning from text. Oral reading, on the other hand, may be undertaken for the purpose of producing expressive oral text. The traditional perspective has been that oral reading should give way to silent reading after the early years of schooling (Taylor and Connor, 1982; Holmes and Allison, 1985). Oral reading is believed by many to be a more complex and difficult process than silent reading (Smith, 1982).

Johnson (1982) examined comprehension of text through silent and oral reading with students aged 7 to 9 years. Literal recall was the only variable measured in the study.

No significant difference between the two forms of reading was found for girls. For boys, oral reading was found to produce higher comprehension in younger subjects. This difference disappeared among the older subjects. As a result, no true difference could be confirmed between oral and silent reading modes.

The results of research on silent versus oral reading have been mixed. Juel and Holmes (1981) examined student comprehension and the reading of sentences both silently and orally. The sample included 48 students in grades two and five, divided equally by both grade and reading ability. Comprehension was measured by having students select a picture which illustrated the concept in a short sentence which was either read orally or silently. There was no difference between the reading modes found for good readers. Poor readers were found to spend significantly more time processing abstract words when reading orally than silently, without a change in comprehension. This finding supports the contention of Smith (1982) above.

Holmes and Allison (1985) studied the relationship between four reading modes (silent, silent while listening, oral to self and oral with an audience) and reading comprehension with 48 grade 5 students. Equal numbers of good, average and poor readers were assigned to each group, as were boys and girls. The sample was screened for IQ, and the experiment was controlled for prior knowledge by using unfamiliar material. The method consisted of having each

group work through a package of reading materials in its assigned reading mode, after which students answered comprehension questions. The questions used were divided into literal, inferential and application types.

The study found that silent reading while listening to the text being read produced significantly superior comprehension results for literal and inferential questions, while oral reading to an audience produced significantly superior results for application comprehension. Among reading ability groups, good readers performed significantly worse after silent reading while listening than in the other modes. No significant relationship was found for the average and poor reading ability groups.

An important finding of this study was the lack of significant difference in comprehension after silent reading to oneself and oral reading to oneself. The authors conclude that oral reading to oneself may be as beneficial for developing comprehension as silent reading.

The findings of Holmes and Allison were partially supported by Miller and Smith (1984). They examined the relationship between literal and inferential comprehension after reading orally and silently. The sample consisted of 94 students in six classes ranging between grades 2 and 5. Students were given reading tasks selected on the basis of familiarity and student interest. Post-testing was done using questions prepared by the researchers.

The study found that students reading silently performed significantly better than oral readers on inferential questions, while oral readers outperformed silent readers on literal questions. This differed from Holmes and Allison, who found no difference between silent and oral reading. In terms of ability, low achievers performed significantly better after reading orally, high achievers after reading silently. This is in agreement with Holmes and Allison. Low achievers did best on inferential questions, while high achievers were significantly better with literal questions.

From the above, it would appear that the following generalizations can be made with respect to silent reading and reading comprehension:

- (i) increased time spent reading, especially silent reading, increases student comprehension levels;
- (ii) silent reading is particularly effective with high reading ability learners;
- (iii) silent reading may not be the best method of developing comprehension with young children, especially boys;
- (iv) adding listening to silent reading of texts may be beneficial for average and poor reading ability learners.

Uninterrupted sustained silent reading. A popular means of incorporating silent reading into the school curriculum has been the adoption of Uninterrupted Sustained Silent Reading programs (USSR or SSR). USSR is a planned program of reading practice, not a form of reading instruction. Students in an SSR program are given a fixed

amount of time each day to silently read materials of their own choice. Its chief attributes are (i) no interruptions, (ii) everyone, including the teacher, reads, (iii) students choose their own reading materials, (iv) there are no reports, records or questions asked, (v) a wide variety of materials are available for students to choose from. SSR programs are considered by their advocates to be a form of whole language exercise which provide an opportunity for students to apply reading sub-skills to the whole process of reading (Manning-Dowd, 1985).

Schaudt (1983) reviewed four research studies contained in the Educational Resources Information Center (ERIC) system related to USSR programs. One study reported gains in both reading interest and achievement among grade 5 and 6 students following a USSR program of unspecified length. A second found significant gains in reading achievement with third grade students, but a lack of significant gains from the second and fourth grade students studied concurrently. The other two examined found no significant difference in either reading achievement or attitudes towards reading between students in USSR programs and students not involved.

Manning-Dowd (1985) reviewed twelve studies involving USSR programs. She identified attitude and achievement as the significant variables requiring examination. Of the twelve studies, seven reported a significant gain in student attitudes towards reading (versus three which found no significant difference and two which did not report on

attitude). This was taken as an indication that SSR programs had a strong impact on student reading attitudes.

The results of Manning-Dowd's review with respect to achievement were not as definitive as the attitude measure. Six of twelve studies found significant gains in reading achievement for SSR participants, but five studies reported no significant gains (one study did not report on achievement). In addition, Manning-Dowd found that in most of the studies reviewed the introduction of the SSR program for the experimental group was not paralleled by the provision of an alternative reading program for the control group. In light of the findings of Leinhardt et al. (1981), it may be that additional silent reading time is responsible for the gains in comprehension, and that if the control groups had received an alternate form of reading instruction or practice the significant gains may not have arisen.

From the conclusions of Schaudt and Manning-Dowd it can be seen that sustained silent reading programs have a positive effect on learner attitudes towards reading. The results with respect to any effect on reading comprehension or achievement are inconclusive, however. It may not be realistic to expect gains in reading achievement on the basis of student participation in SSR programs until more definitive studies have been carried out.

Computers and Reading

Computers have been used in reading instruction for a number of years. Most computer reading programs are based on the reductionist reading approach, following Skinnerian programmed instructional models to develop reading sub-skills in students (Miller, 1984; Balajthy 1987; Geoffrion and Geoffrion, 1984/85). CAI in reading comprehension tends to involve learning in letter and word recognition, vocabulary development and elementary comprehension level only. Context is rarely a part of CAI in reading programs (Miller, 1984). The CAI exercises are usually direct transcriptions of workbook drill and practice exercises, with immediate feedback the only computer based enhancement (Geoffrion and Geoffrion, 1984/85). Like SSR, CAI in reading exercises provide students with an opportunity to practice learned skills. They do not usually directly teach new material (Kinzer, 1986).

CAI exercises which deal with reading comprehension passages tend to ask questions which are convenient to ask, rather than questions which will assist in the development of student comprehension of text. They rarely make use of students' schemata, as one cannot tell if students are answering questions based on understanding the reading or simply from prior knowledge (Geoffrion and Geoffrion, 1984/85). As a result, comprehension CAI programs steer away from subject matter students are familiar with. This reduces the

affect of both schemata and interest, powerful predictors of success in reading comprehension.

The result of this picture of CAI in reading has been considerable criticism of its usefulness. Blair, Rupley and Jones (1986) examined opinion and research in computers and reading instruction, and found that few studies of the effects of computer based learning have been conducted, and of those that exist most were undertaken using mainframe computers prior to the development of the microcomputer. This echoes the concerns of Bass, Ries and Sharpe (1986).

One frequently cited reason for using computers in education has been their motivational value (Parry et al., 1985/86; Becker, 1984; Balajthy, 1987). There is currently no clear understanding of whether this motivational value arises from something intrinsic about computers, or is simply another manifestation of the Hawthorne effect (novelty). There is some concern that as computer use becomes commonplace, many of the effects found by researchers may disappear, as they were a result of novelty rather than some deeper motivation in students to learn using the computer (Balajthy, 1987; Thompson, 1980; Blair, Rupley and Jones, 1986).

Pedagogy. A second concern of Blair, Rupley and Jones was the lack of pedagogical understanding of how computers should be used in classrooms. They suggest that microcomputer CAI in reading is rarely co-ordinated with the regular instructional program in any real way. Students

engage in games and drill with no connection to the reading program. As a result, the affect of CAI in reading in improving student reading ability is minimized.

Winkler, Shavelson, Stasz, Robyn, and Feibel (1985) examined the factors which are involved in the use of computers in classrooms, and suggested that "sound classroom microcomputer use will occur when teachers make reasonable decisions about matching the microcomputer and available courseware to the instructional goals, the structure of the subject matter, the nature of the students, and the context of instruction" (p. 287). These are the factors which Blair, Rupley and Jones (1986) claim are lacking in classrooms attempting to use microcomputers in reading instruction today.

The belief that computer use must be integrated into regular classroom activities is also put forward by Schiffman (1986). She uses the term software infusion to describe a situation where appropriate computer software is used to achieve goals in curricular instruction which are best achieved through the use of the computer.

Computer delivery of text. Computers deliver text on a monitor screen, no more than a screenful at a time. It is possible that this variation from the standard book style of delivery may have an impact on reader comprehension.

Gerrell and Mason (1984) examined the impact of computer controlled text on a monitor screen to see if it has the same impact as printed text on students. The sample

involved 30 grade five students teacher-nominated for average to above average reading ability. The students read two portions of the same text, one on a computer, the other on paper, and answered comprehension questions. It was found that student comprehension of the computer chunked text was superior to that of the paper material.

Gerrell and Mason failed to take into consideration the Hawthorne effect in their conclusions. This was a short term study, and there is a very significant chance that superior comprehension arises more from novelty than from actual superior learning. It is safe to conclude from this, however, that text on the computer screen does not have any seriously debilitating characteristics for students.

Research on CAI and achievement. A significant amount of research has been conducted since the 1960's on the relationship between all forms of computer assisted instruction and student achievement. Nierniec, Samson, Weinstein and Walberg (1987) reviewed 48 studies of CAI and student achievement in elementary schools. They found that CAI has a significant effect on achievement. CAI resulted in an effect size² of .45 standard deviations for CAI students beyond their control groups. In addition, they found that high achievers gained less than low achievers, younger learners gained more than older learners, and that the highest effect size occurred with CAI drill exercises. It was also found that the effect size of published material was significantly greater than that of unpublished dissertations.

In concluding, the authors suggested that CAI may be best used with lower order achievement tasks, such as paired association. Computers were found to enhance achievement when the learning task was relatively simple. The smallest effect size (.11) was found with reading comprehension, a complex learning.

The results of the above study were echoed in earlier works by Samson, Niemiec, Weinstein and Walberg (1985) and Kulik, Bangert and Williams (1983), who independently conducted meta-analyses of research on CAI and student achievement at the secondary level. The major difference was a smaller generalized effect size found (.32 compared with .45). The study by Samson et al. replicated the findings with respect to the effects of CAI drill and practice, and the greater effect size of published materials.

These studies are not without their critics, however. Clark and Leonard (1985) found that many research studies in CAI have been critically flawed, either through lack of random assignment to groups or control groups that do not receive equivalent instruction to experimental groups. They examined the works of Kulik in the early 1980's. After eliminating those studies that were biased as described above, only two of fifteen reported significant findings in favour of CAI. These findings are consistent with earlier findings of Clark (1983) and Jameson, Suppes and Wells (1974).

Even with the concerns noted above, there seems to be sufficient evidence to accept that computer based learning is

an acceptable form of learning for students. Parry et al. examined syntheses of research relating computer based learning and achievement, and concluded that CBL (i) should be used as a supplement, not a replacement for regular classroom instruction, (ii) is probably more motivating than traditional classroom methods, (iii) can save time in learning, and (iv) is especially effective with low achievers. These conclusions make no claim about the superiority of computer related learning. They are also related to achievement in general, not to reading achievement.

Research on CAI and reading achievement. Ten studies relating CAI in reading and reading achievement were obtained through a manual search of Dissertation Abstracts International and a computer search of the ERIC database. All involved drill and practice and tutorial forms of CAI. Of these, two (Litman, 1978; Ortmann, 1984) reported a significant positive finding in favour of CAI. The others (Anelli, 1978; Easterling, 1983; Porinchak, 1984; Bryg, 1985; Coomes, 1986; Cooperman, 1985; Bass, Ries and Sharpe, 1986; Icabone and Hannaford, 1986) found no significant difference between the control and experimental groups after periods of CAI in reading ranging from fifteen weeks to one year. This cursory review therefore does not support the effectiveness of CAI in reading in improving student reading comprehension beyond that of regular classroom instruction.

This conclusion was supported by the work of Thompson (1984), who conducted a similar non-systematic review of

papers and dissertations dealing with CAI in reading and reading achievement. All studies utilized programmed instructional model CAI in reading. Thompson concluded that CAI was an effective means of developing student reading achievement, but not more effective than traditional instructional methods. He also found that students lost interest in poorly constructed CAI just as quickly as in poorly constructed classroom lessons.

This view is by no means conclusive. Castner (1982) conducted a non-quantitative review of sixteen studies comparing the effectiveness of drill and practice and tutorial CAI with paper based drill and practice exercises in reading, writing, spelling and vocabulary. The conclusion was that in the drill and practice environment, CAI produced superior results to paper based methods. No rationale for the decision other than a comparison of study findings was given.

Summary of computers in reading instruction. With the exception of Castner's conclusions, it would appear that CAI cannot be said to produce higher levels of reading achievement than traditional instruction. This is not to say that CAI in reading has no value: It simply means that working at sub-skills with a computer is an alternative that teachers can make use of, confident that students so engaged will achieve as well as students who practice their sub-skills in other fashions. Since there seems to be a relationship between the acquisition of reading skills and cognitive style (Pitts and Thompson, 1982), it may be that particular

students would benefit more than others on a similar basis. No research was found on this topic.

A second limitation on the use of CAI in reading is that all the research reviewed above is based on CAI drill and practice and tutorial exercises that reflect the reductionist model of reading instruction. While the whole language model does not totally reject what is learned in sub-skills, it considers such learning to be lacking in real meaning for learners because exercising a particular sub-skill or set of sub-skills does not reflect the real process of reading.

According to Miller (1984), the whole language approach must involve the computer through an emphasis on process, not product. Students must be able to read whole texts for meaning, not simply drill on individual portions of the reading process. Language must be presented in context to the learner, preferably involving as many parts of language arts as possible (reading, writing, viewing, speaking, listening), woven together throughout the learning.

Despite the results which show that CAI is valuable in the development of reading achievement, some theorists remain skeptical. Thompson (1983) indicated a concern with the pre-determined structure of computer programs, and the possibility that they may force students into restrictive, pre-selected technical paradigms of reading which may foster mechanistic styles of thought. In Thompson's view the goal of reading instruction and practice is reading itself, and at some point students must leave the computer and read

printed materials. He asserts that unless computer based learning can foster the interest and ability to read non-computer oriented materials, computer related reading instruction is not valuable.

Computer Based Interactive Fiction

There are many who believe that computer based interactive fiction can satisfy the needs of both Miller and Thompson. Interactive fiction has existed for many years in book form, but the books tend to have very short story lines due to the need for multiple plots and branching. As a result there is less character and dramatic development than in most non-interactive stories (Packard, 1987).

Computer based interactive fiction breaks through these limitations. Huge amounts of text can be stored on a single computer disk, and the branching and multiple plots are totally transparent to the reader. Long, complex stories which take weeks to complete exist as computer text adventures.

A key element in all recommendations for using text adventures as reading materials is the degree to which students are motivated to play them. There may be several explanations for this motivation. They are personally involved in the story, and actively shape what happens as the plot proceeds (Grabe and Grabe, 1985; Little, 1985; Brady, 1986). The computer itself may draw some students to read text who might otherwise not have been interested (Layton,

1987), although this may be a reflection of the Hawthorne effect. Whatever the reason, students given the opportunity to play text adventures tend to be very motivated and excited about the reading (Kuechle, 1983; Little, 1985; Jarchow and Montgomery, 1985; Brady, 1986; Williams, 1984). Students are frequently motivated to read text adventures which are significantly above the reading level at which they are comfortable with paper based stories (Jarchow and Montgomery, 1985).

The benefits of reading text adventures have been postulated by a number of proponents of the whole language approach to reading. Layton (1987) identified fifteen attributes of reading and thinking which are exercised through playing adventure games, including recognizing and recalling important details, drawing inferences, recognizing and understanding figurative language, skimming and scanning, generalizing from facts, learning content-specific vocabulary, and understanding synonyms and synonymous phrases. All of these attributes are developed in a holistic fashion rather than discretely as sub-skills.

Most of the whole language theorists point to the need for active comprehension of the text as the key to developing reading skills while playing text adventures. In order to make the text adventure proceed, the reader must understand the text presented and respond appropriately. Errors in comprehending the text result in poor responses from the reader, which cause unexpected and undesirable

changes in the plot of the adventure (Brady, 1986; Grabe and Grabe, 1985; Kinzer, 1986; Layton, 1987). Playing adventure games may also enhance the understanding of text structure and story grammars (Anderson-Inman, 1987), which may in turn increase reading comprehension. Role playing has even been suggested as a means of developing schemata for situations which fall outside the bounds of normal modern experience (Mavrogenes, 1983).

One study involving computer based interactive fiction and reading was located after extensive searching of the ERIC database. Lancy and Hayes (1986) examined the impact of playing adventure games on student interest in reading. The study, undertaken over a four week period, involved a sample of eight students at a summer workshop who ranged between grades 5 and 9.

The subjects were pre-tested with the Wise Reading Attitude Inventory, which showed that none of them had more than average reading interest. The students were then encouraged to play text adventure games for up to three hours per day, four days a week for four weeks. Encouragement in this case meant that students were given clues, but no reading help with the text adventures. Student progress and interest were monitored by observation throughout the experimental period. No post-testing of any kind was undertaken.

The researchers found that the students were willing to play the games for extended periods of time without any

pressure of any kind. The only problem that arose was when the student stalled in the game, not knowing how to proceed. This was corrected by clue giving. From this the researchers concluded that students with no more than average interest will spend extremely long periods of time playing adventure games, and they recommend permitting the practice in classrooms as a means of encouraging the development of reading comprehension.

The lack of any kind of post-testing in this study makes drawing any conclusion difficult. The researchers did not take into consideration the potential of the Hawthorne effect in their study. Also, it is possible that the sample was biased, as this exercise occurred during the summer, and the subjects may have been volunteers. A final concern is that the sample was so small that it is not possible to generalize the finding, such as it is, to the general population.

The study reviewed above points to the need for an empirical examination of the relationship between student reading of computer based interactive fiction and the development of reading comprehension.

Summary

This review has attempted to survey the literature pertinent to reading comprehension, computer based reading instruction and computer based interactive fiction as it relates to the study being undertaken. It was seen that the major predictors of student reading comprehension are the

availability of schemata, knowledge of story grammars and personal interest. The most significant instructional factors in the development of reading comprehension are the amount of time spent reading each day, especially silent reading, and the addition of concurrent listening with silent reading for students of poor and average reading ability.

Uninterrupted sustained silent reading programs, a popular means of increasing silent reading time in the classroom, were found to increase student interest in reading. No definite effect on reading comprehension was discovered, however.

Computer based learning in reading instruction was found to fall largely into the realm of the reductionist model of reading. Nearly all CAI in reading programs employ Skinnerian, programmed instruction models in their drill and practice and tutorial methods. While research tends to support the effectiveness of these types of programs generally, little evidence was found to suggest that CAI in reading is more effective than classroom instruction in developing reading achievement. It was found to be equivalent to traditional instructional methods in terms of student learning outcomes, however.

Existing CAI in reading is not structured in a fashion that is acceptable to proponents of the whole language model, who require students to work with full texts and read for the meaning of the text. Computer based interactive fiction has been suggested as an alternative to CAI in reading which

satisfies the whole language model, and may bring with it powerful motivational and instructional factors which may, in turn, foster the development of student reading comprehension. It is the purpose of the present study to examine this potential.

In reviewing the research studies above, a number of concerns were found with the methodologies which may have created biases in the findings. The most significant were:

- (i) failure to conduct the experiment within a real instructional environment;
- (ii) failure to randomize within the subject groups, or the use of volunteer subjects;
- (iii) failure to provide equivalent instruction to the control group;
- (iv) failure to take the Hawthorne effect (novelty) into consideration; and
- (v) failure to balance observational data with post-testing of subjects.

The present study will attempt to deal with each of these problems in its experimental design.

Endnotes

1. It is through schemata that a reader can use the context of a story to comprehend multiple word meanings. For example, the juxtaposition of the words 'bark' and 'canoe' invoke the schemata related to both words, and the reader would identify the correct relationship as tree bark. A person who does not know anything about birch bark canoes (therefore lacking the appropriate schema) might have difficulty comprehending text containing these relationships.
2. Effect size is a measure of the effect that a treatment can be expected to have on a population, based on experimental results. It is given by the equation:

$$\text{Effect Size} = \frac{\text{Mean (Experimental)} - \text{Mean (Control)}}{\text{Standard Deviation (Control)}}$$

and is used in order to compare the results of dissimilar studies (Niemi et al., 1987).

Chapter 3 -- Experimental Design

The experiment upon which this study was based was designed to minimize or eliminate as many of the concerns with research studies that were discussed in the last chapter as possible. In keeping with Bass, Ries and Sharpe (1986), the experiment involved whole classes of students in an actual school setting, using microcomputers and commercially available software. In addition, all of the testing instruments used in the study were obtained from the local school district, and are available commercially.

The subjects involved in the sample were selected on a non-volunteer basis, although parental approval was sought and received prior to the beginning of the experiment. Assignment to groups was done on a strictly random basis. Reading instruction, in the form of USSR, was provided for the control group in equal duration to the experimental group. The dependent variable, reading comprehension gains, was a measured rather than strictly an observational variable, through both pre- and post-tests.

The only concern found in Chapter 2 not addressed directly in this study was the Hawthorne effect. It was partially compensated for in that all students involved had received regular CAI drill and practice in language and mathematics skills all year long, plus some introductory word

processing instruction. Computers were therefore not new to any of the students.

Design

The experiment followed a true experimental pre-test post-test comparison group design after McMillan and Schumacher (1984). The sample subjects were assigned to experimental and control groups on a purely random basis (random numbers selected from a table and applied to class list position). All students were pre-tested to determine ability, level of achievement and reading comprehension. The treatment, consisting of students playing text adventure games at individual microcomputers, lasted eight weeks. At the conclusion of the treatment period, all students were post-tested using a second form of the same test used in the pre-test of reading comprehension.

Background

The site of the study was a public school in a northern community of 35,000 people. It was situated in an affluent subdivision, in which most of the residents were owners of their homes. The school's population during the 1986/87 school year was approximately 450 students in Kindergarten to grade 8. The entire upper-elementary section of the school (grades 4, 5 and 6) had adopted a thematic approach to instruction, including the whole language model of

language learning. As a component of this, all students engaged in 15 minutes of USSR daily.

Sample

The subjects that made up the sample in this study were grade five students. Two full classes of students were selected, a total of 50 students. Of these, one did not receive parental consent to participate, one moved away during the study, and three were dropped from the final count due to incomplete testing. The result was a sample of 45 students, 26 males and 19 females, ranging in age from 10 years 1 month to 12 years 6 months.

Socio-economic and personal data were not gathered relative to the family backgrounds of the students involved in the sample, as the school's community was relatively homogeneous.

Method

The experimental treatment involved students engaging in computer based interactive fiction through playing text adventure games on an Apple //e microcomputer for 15 minutes per day, every day for eight weeks. The comparison treatment, which the control group followed, involved uninterrupted sustained silent reading at the same time as the experimental group was engaged in interactive fiction.

Students in the experimental group were allowed to choose from among four different text adventure games, all

produced by the Infocom Corporation (MoonMist, Wishbringer, Planetfall and Seastalker). The text adventures produced by the Infocom Corporation were selected because they all are purely text adventures, using no computer graphics of any kind. In addition, they are integrated packages, containing information useful in the solution of the game in the form of photographs, letters, and relevant maps. Infocom games are rated for difficulty by the producer, unlike many adventure games. All of the games used in this study were rated for beginners.

Within a few days, all students settled on either Wishbringer or MoonMist. Students in the control group were allowed to read anything they wanted, either brought from home, or chosen from library or classroom materials at the school. Their selections ranged from novels to comic books.

The students involved in the sample were placed in experimental and control groups through random number assignment. The only control on selection was that half of each group came from each of the two grade five classes that made up the sample of the study.

Treatment. Following the reading comprehension pre-tests, the students in the experimental group began playing adventure games in the school computer lab. All treatments occurred within a single half-hour block in the morning of each school day. As the lab could only accommodate one-half of the experimental group at one time, the experimental group was divided into two halves, by homeroom class, and

each half worked with the computers in turn. The second group always followed the first group immediately.

Playing text adventure games on a computer required the students involved to read text that appeared on the monitor of the computer, then respond to it by typing in a short sentence or phrase which the computer accepted as a command. This might involve such actions as 'examine the lamp', 'go up the stairs', 'look under the bed', 'search under the carpet', or 'wait'. The student must understand what has happened to the point where they are in the story, and must respond with some reasonable action that they would like to take. If the student does not type anything, the adventure game sits and waits: No prompting or computer-directed action takes place. If the student types in a command that cannot work in a given situation, the computer replies that the action is not possible ('You cannot go north from here') or asks for clarification ('You have to tell me how to do that to a door'). Absurd commands are either turned down flat ('You cannot walk on water!') or responded to in a mildly sarcastic manner ('You raise your fist to strike, but then decide, Aw, it's not a bad cat anyway, and leave the cat alone').

Time on the computers was strictly limited to the fifteen minute time slot allotted for USSR. The text adventures used in the study permitted the players to save their games and return to the same spot in the next session. Students made their way to and from the lab on class time

outside the fifteen minute reading time. This required only one minute in either direction, and so did not pose a problem for the teachers left behind with the control groups, who engaged in USSR while the experimental group was away.

Instrumentation

In order to ensure that this experiment would be easily replicable, a major condition relative to the selection of testing instruments was that all tests used must be commercially available, well known and widely used in the testing field at large.

Descriptive pre-testing. Prior to undertaking the study, information about the achievement and ability of the sample population was obtained through the use of commercial standardized tests. The local school district had conducted standardized achievement tests of all students in grades 2, 5 and 8 in January of 1987, using the Canadian Test of Basic Skills (CTBS) (King, 1981). The results of those tests had been machine scored and were available for all students in the sample. Vocabulary, Reading, Spelling, Language and Overall Achievement data were obtained from this source. The data for each student had been converted from raw scores into Grade Equivalents by the scoring service.

Student ability testing was not conducted in the same manner by the school district, and therefore was undertaken for experimental purposes. After consultations with school district officials, it was determined that the Canadian

Cognitive Abilities Test (CCAT) (Wright, 1974) would be an appropriate measure. The verbal and non-verbal batteries of the test were administered to all students participating in the sample shortly after the beginning of the experimental period. The results of these batteries were hand scored using masks available from the publisher. Raw scores were converted into Standard Age Scores (IQ equivalents) through tables provided by the publisher (Wright, 1974). The mathematical battery was not administered as mathematical aptitude was not considered a relevant variable in this study, and because the mathematical battery of the CCAT was perceived by local officials to be confusing for students.

Reading comprehension testing. In consultation with local school district officials, it was decided that pre- and post-testing for student reading comprehension would be done through use of the Stanford Diagnostic Reading Test (SDRT) (Karlsen, Madden and Gardner, 1976). This test was selected on the basis of its longevity in the reading assessment field, the availability of alternate test forms of known reliability (mean co-efficient of reliability of all sub-tests 0.833), and its local availability in the school district.

Form A of the SDRT Brown Level was administered as the pre-test to the sample under identical conditions, simultaneously in their homeroom classes. Students who were absent for the test were tested immediately upon their return to school. Form B of the same level was used as the post-test. All tests were hand-scored using masks produced

by the test publisher. Both forms included sub-tests in Phonetic Analysis¹, Structural Analysis², Auditory Vocabulary³, Literal Comprehension⁴, Inferential Comprehension⁵, and Reading Rate⁶. All sub-tests were completed. Raw scores from all tests were converted into Scaled Scores (permitting between-form comparisons) using tables provided by the publisher (Karlsen, Madden and Gardner, 1976).

Other measures. Student age was calculated as of April 1, 1987 using the school's official register for the two classes involved. Student attendance during the experimental period was a possible bias in the results, and therefore was recorded from the school's official register.

The sample students were members of two different grade 5 classes, taught by two different teachers. As a result, homeroom class was recorded as a potential source of bias.

Lastly, the students involved in both the experimental group and the control group selected the materials they wanted to work with, and these varied widely in reading difficulty. A measure of the reading level of the materials read was obtained by taking a random sample of the text materials, and a copy of the interactive fiction used, and typing at least 150 words of each into the MouseWrite word processor on an Apple IIGS microcomputer. The word processor uses a combination of the Frye Graph and Flesch

Index to estimate the reading difficulty of the passage in memory (Cochard, 1986).

Summary

The experiment employed a true experimental pre-test post-test comparison group design. The sample included 45 students in grade 5, split between two classrooms in one school. Experimental and control groups were created through random selection. The treatment had the experimental group play text adventure games while the control group engaged in USSR for comparison purposes.

Measures of student achievement and ability were obtained to describe the sample. Class, age and difficulty of the reading material were recorded for all students. Pre- and post-tests of reading comprehension were undertaken using the Stanford Diagnostic Reading Test. The results of all of these measures are provided in Chapter 4.

Endnotes

- 1 Phonetic analysis refers to the student's ability to identify sounds in words. Testing involves asking students to relate the sound heard with the sound contained in a similar word (Karlsen, Madden and Gardner, 1976).
- 2 Structural analysis refers to the ability to decode words through the analysis of word parts, including syllables, prefixes, suffixes and root words (Karlsen, Madden and Gardner, 1976).
- 3 Auditory vocabulary tests the word knowledge of students through having them identify words which are defined orally, thus not involving their ability to read (Karlsen, Madden and Gardner, 1976).

- 4 Literal comprehension refers to ideas and information which are explicitly given in the text which is read. Students are tested on recognition and recall of such facts (Pumfrey, 1977).
- 5 Inferential comprehension refers to a student's ability to use information obtained through reading to conjecture and hypothesize. It is significantly involved in decision making (Pumfrey, 1977).
- 6 Reading rate refers to the ability of students to read quickly and with comprehension passages which are considered simple. The testing procedure involves a Cloze test (Karlsen, Madden and Gardner, 1976).

Chapter 4 -- Analysis of the Data

Introduction

All test results obtained from the experiment were keyed into a spreadsheet for manipulation. This yielded means, standard deviations, and permitted the sorting of the data by variable in order to obtain variable-specific results. Sorted data was transferred to the StatView statistics package on a Macintosh microcomputer, where skewedness, correlation, multiple regression, t-tests and ANOVA were calculated, and graphs generated.

A first look at the pre- to post-test gains results reveals extraordinarily large standard deviations compared with means. This is in part due to the conversion of raw scores with values of less than one hundred into scaled scores which ranged from 300 to 900. The effect of this conversion was to enlarge the apparent magnitude of the standard deviations without increasing the value of the mean gains. The issue of the quality of the data will be dealt with as part of the discussion of the findings of the hypotheses in Chapter 5.

Sample Characteristics

The statistical characteristics of the total sample are given in Table 1. It provides comparative data for both the

experimental and control groups in terms of ability, achievement and other characteristics.

Ability. The students participating in the sample were found to be statistically equivalent in verbal ability and overall ability. However, a slightly significant difference was found for non-verbal ability (control group higher than experimental, $t = -1.404$, $p < 0.1$). Further examination of means and skewedness reveals a non-significant difference

Table 1
Comparison of Population Data between
Experimental and Control Groups

Ability Data	n =	Experimental Group		Control Group		Unpaired t-value
		22		23		
		Mean	Skew	Mean	Skew	
CCAT Verbal Test		104.545	.215	99.913	-1.155	.946
CCAT Non-Verbal Test		95.636	-.43	102.261	-.791	-1.404*
Average		100.091	-.188	101.087	-1.071	-0.223
Achievement Data						
CTBS Vocabulary		5.550	-.026	5.391	-.808	0.443
CTBS Reading		5.582	.328	5.574	-.514	0.250
CTBS Spelling		5.659	-.482	5.391	.212	0.651
CTBS Language		5.614	-.58	5.635	-.392	-0.075
Comprehensive		5.577	.082	5.591	-.775	-0.051
Other Variables						
Sex (1=Male, 2=Female)		1.455	.183	1.391	.445	.421
Age		10.8	.592	10.93	1.187	-0.807
Experimental Conditions						
Text Difficulty		5.45	-1.021	3.62	1.64	4.244**
Attendance		35.773	-1.056	36.043	-1.555	0.353

* $p \leq 0.1$

** $p \leq 0.05$

in verbal ability in favour of the experimental group. The verbal and non-verbal differences balance out in the ability average. In addition, inspection of the non-verbal ability test showed that it was not related to skills or abilities in reading or language at all. For these reasons, average ability was chosen as an appropriate independent variable to be examined relative to reading comprehension in the findings.

Achievement. The figures in Table 1 show that there was no statistically significant difference between the experimental and control samples in any of the achievement measures taken by the Canadian Tests of Basic Skills. As with ability, however, an examination of skewedness reveals two large if statistically insignificant differences in several of the variables between the two groups. The differences are found in reading achievement (more high ability students in the experimental group, low in the control), and spelling achievement (low in the experimental group, high in the control group).

Others. Measures of the distribution of boys and girls, and the occurrence of absences during the experimental period were also taken. Neither Sex nor Attendance was found through Table 1 to be either significantly different or skewed differently between the experimental and control groups.

On the basis of the results listed above, the experimental and control groups were considered to be equivalent for the purposes of this study.

Experimental Conditions

Values obtained for Text Difficulty and Attendance are also presented in Table 1. A significant difference ($t = 4.244$, $p \leq .05$) was found for text difficulty. The text adventures used by students in the experimental group tended to be higher difficulty reading materials than those read by students of the control group. Attendance was neither significantly different nor skewed between the experimental and control groups.

Other experimental condition variables were controlled through the nature of the experimental design. These included homeroom class, which was controlled through equal random assignment to the experimental and control groups, testing conditions (all tests were written concurrently), time conditions (the treatment occurred within the same half hour block with the two groups each day), and maturation (controlled through the eight week timeframe of the experiment).

A potential threat to the validity of the experiment was the attitude of the students who were assigned to the control group. The treatment was considered 'fun' by the students, and there was a possibility of disappointment at not being selected affecting the performance of some students as they completed the necessary tests. An attempt was made to control this by promising to let the control group participate in the treatment once the experimental period was over and all tests were completed. The foregoing analysis of the ability

and achievement statistics suggests that this did not result in a problem in the data.

Findings of the Hypotheses

The primary results related to the hypotheses are presented in Table 2. The minimum acceptable criterion for significance in T-tests and ANOVA's was established as the ninety percent confidence limit ($p \leq 0.1$).

Hypothesis 1. The first null hypothesis predicted no difference in the gains in overall reading comprehension between the experimental and control groups. Examination of the data shows that the difference between the groups was insignificant ($t = -0.684$). This lack of difference is shown in the frequency distributions of Figure 1. The non-significant

Table 2
Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills
between Experimental and Control Groups

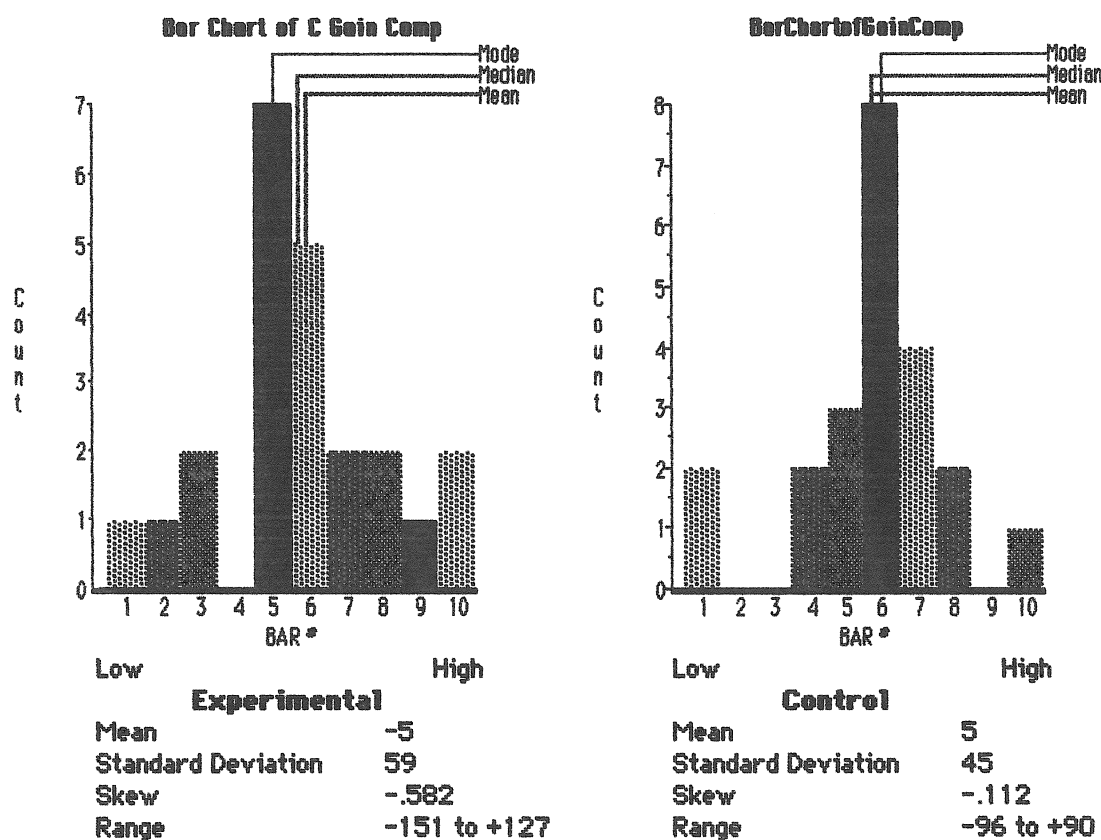
<u>n =</u> <u>Reading Component</u>	Experimental 22			Control 23			Unpaired t-value
	\bar{X}	σ	Skew	\bar{X}	σ	Skew	
Total Comprehension	-5	59	-1.302	5	45	-.68	-0.684
Phonetic Analysis	-26	96	1.331	-12	61	.212	-0.589
Structural Analysis	28	47	-.554	10	39	-.429	1.431 *
Auditory Vocabulary	-12	89	-.549	-23	69	.082	0.49
Literal Comprehension	-9	52	.159	2	59	-.083	-0.664
Inferential Comp.	0	70	-.582	7	55	-.112	-0.331

* $p \leq 0.1$
 ** $p \leq 0.05$
 *** $p \leq 0.01$
 **** $p \leq 0.005$

Key: \bar{X} group mean
 σ standard deviation

difference between the groups reflects greater mean gains for the control group than the experimental group. As a result, the decision was to accept the first null hypothesis.

Figure 1
Frequency Distributions of Gain Scores in Total Comprehension
for Experimental and Control Groups



Hypothesis 2. The test data related to the second hypothesis are also included in Table 2. The second null hypothesis stated that there would be no significant difference between the gains of the experimental and control groups in any of the sub-skills measured by the SDRT.

The Stanford Diagnostic Reading Test examined six sub-skills in analyzing reading comprehension. Total reading comprehension was measured as the combination of literal

comprehension and inferential comprehension scores. In addition, phonetic analysis, structural analysis and auditory vocabulary were tested. The sixth, reading rate, was not included in this study.

The results in Table 2 indicate that a significant difference existed in one of the five sub-skills of reading comprehension. The mean gain in structural analysis was significantly higher ($t = 1.431$, $p \leq 0.1$) than that of the control group. Of the other four sub-skills, the mean gains of three (phonetic analysis, literal comprehension and inferential comprehension) were higher for the control group than the experimental group. Only auditory vocabulary exhibited a higher mean gain for the experimental group than the control group. None of these differences in means were significant, however.

On the basis of these findings, the second null hypothesis was rejected.

Hypothesis 3. The results related to the third and fourth hypotheses are contained in Table 3. The third hypothesis predicted that there would be no difference between students of high, average and low ability in terms of their gains in overall reading comprehension within the experimental group. For this test, the experimental groups were divided by average ability as found on the CCAT. Divisions by groups were made on the basis of one population standard deviation from the population mean (100 ± 15 , as the scores were scaled from the manual). Analysis of Variance

across the three resulting groups yielded an insignificant result ($F(2,21) = 1.128$). A further t-test analysis failed to discover any within-groups significant difference that might have failed to show up on the ANOVA. As a result, the decision was to accept the third null hypothesis.

Table 3
Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Ability Groups

Experimental Group	n =	Ability (CCAT Avg)			t-tests (Unpaired)			
		High	Average	Low				
Reading Component		Mean	Mean	Mean	Hi:Avg	Avg:Low	Hi:Low	F ratio
		8	10	4				
Total Comprehension	23	-21	8		1.315	-0.9	0.408	1.128
Phonetic Analysis	-69	-11	-21		-1.038	0.252	-0.561	0.643
Structural Analysis	51	28	0		0.89	1.282	1.468*	1.379
Auditory Vocabulary	-58	-5	25		-1.076	-0.588	-1.646*	1.064
Literal Comprehension	-22	-15	28		-0.261	-1.421*	-1.831*	1.343
Inferential Comp.	56	-16	-16		1.913**	0.013	1.84*	2.279

Control Group		High	Average	Low	Avg:Low t-test
Reading Component	n =	1	18	4	
		Mean	Mean	Mean	
Total Comprehension		53	11	-31	1.872**
Phonetic Analysis		-68	-15	13	0.799
Structural Analysis		26	10	8	0.066
Auditory Vocabulary		26	-38	30	-1.847**
Literal Comprehension		66	-1	2	-0.094
Inferential Comp.		44	19	-57	2.875****

*	p ≤ 0.1
**	p ≤ 0.05
***	p ≤ 0.01
****	p ≤ 0.005

Hypothesis 4. The fourth hypothesis expected no significant difference in mean gains between ability groups on any of the reading comprehension sub-skills. Examination of the results in Table 3 shows that a clear acceptance or

rejection of the hypothesis cannot be made. The results of the ANOVAs indicate that no overall significant difference between the experimental and control groups existed for any of the five sub-skills. However, t-tests performed within-groups showed a number of very strong differences, chiefly in but not restricted to the high-to-low comparisons. Only structural analysis, in which the entire experimental group differed from the control group significantly, lacked a within-groups significant result.

Comparison with the results for the control group provides some clarification. The control group high ability 'group' consisted of one subject, which precluded any within-groups testing or ANOVA tests. The average-to-low within-groups tests revealed that for two of the five, significant differences by ability existed. This indicates that any within-groups differences in the experimental group were probably a result of ability rather than the treatment involved. Consequently, the fourth null hypothesis was tentatively accepted.

The decisions relative to the findings of the hypotheses are summarized in Table 4 below.

Discussion of the Findings

Observational data. During the treatment period, observations of the experimental group were logged by the author, who acted as the supervising teacher. Several of these observations may bear a relationship to the results.

Table 4
Summary of Hypotheses and Decisions

Hypothesis	Decision	Basis
1 Experimental group mean gains in reading comprehension no different than gains of control group.	Accept	$t = -0.684$ not significant.
2 Experimental group mean gains in each reading comprehension sub-skill no different than mean gains of control group.	Reject	Structural Analysis $t = 1.431$ ($p \leq .1$)
3 Subjects in experimental group will not differ by ability in mean gains in reading comprehension.	Accept	$F = 1.128$ not significant.
4 Subjects in experimental group will not differ by ability in mean gains in any sub-skill of reading comprehension.	Accept	F ratios not significant.

It was found that lack of knowledge about the way text adventures operate hindered student progress during the early days of the treatment. This included both computer operations and the nature of playing adventure games.

The text adventures used in this experiment permitted the players to quit at any time in the story, and continue where they left off in the next section. The difficulty presented with this was that it required the students to go through a 'SAVE' procedure which involved removing the master disk, inserting a 'save game' disk, and pressing keys to cause the computer to do the saving at the correct moment. Over the first two weeks, many games were lost, and several master disks damaged due to student errors.

The second problem was lack of familiarity with adventure games themselves. While the Infocom games had very well developed parsers¹, the students were not initially well versed in those words that would most efficiently result

in actions in the game. A list of appropriate words was provided initially, and an explanation of how to interact with the game was given, but it took the students the first several days to become comfortable with the genre of their text adventure.

A third factor that delayed the progress of some students was settling on which text adventure game they wanted to play. Four choices were available, and over the first several days a number of students found that for interest or difficulty reasons they wanted to abandon the game they had started and switch to another. This process ended by the end of the second week for all participants.

The result of these three factors was markedly slower progress for the majority of students over the first two weeks of the experiment than the last six, which may have resulted in weaker gain scores than otherwise might have been expected. This may explain the generally wider standard deviations in the gain scores of the experimental group than were found for the control group.

In observing the students it was found that those who had difficulty making progress in the game were the ones who had problems with spelling or vocabulary. The spelling problem resulted in a constant string of messages from the computer that it could not understand the command being typed in. Vocabulary problems were noted when the student could not understand the text on the monitor and had to ask for help. Help in both cases was provided by asking

appropriate questions to aid the student in discovering the correct solution. Only as a last resort would a spelling or word definition be given directly.

Another problem that resulted in slow progress for a few students was related to their problem solving ability. Several times students would request help because they did not know what to do next, and questioning revealed that they had understood the text to that point. They simply were not sure what action to try next. Help in these circumstances amounted to asking questions to assist the student in deciding on a course of action. If this did not elicit a suitable response, a course of action would be suggested. Problem solving ability is a variable that was not measured during this experiment, so its real effect cannot be determined at this time.

The role of student interest in the development of reading skills and the possibility of confounding due to novelty were discussed in Chapter 2. The interest of the experimental group in the treatment was plain for any observer to see throughout the treatment period, and was most noticeable when the experiment drew to a close. It would thus appear that novelty was not a factor in maintaining student interest in the text adventures. Student interest in the control group was student controlled, as they were able to select whatever materials they wanted to read.

Two anecdotes serve to illustrate the impact of interest in the experimental treatment. In the first case, one of the

students in the experimental group had to be dropped from the final data analysis because s/he did not complete any part of the comprehension section of the pre-test. This student was found to have the lowest overall ability rating of the sample group. S/he was a special education student who had been mainstreamed, and was receiving pull-out assistance on a regular basis in mathematics and language arts. Despite intensive problems in getting through the story due to spelling and vocabulary problems, the student persisted daily, and remained extremely interested throughout the period. S/he was a student that had had numerous negative disciplinary contacts with the author in the past. Those negative contacts ended, and were replaced with gifts of art work created by the student. Of greatest significance was the fact that the student completed the entire post-test, and while not showing extremely high scores, did not score at the bottom of the sample either. Something in the experimental environment caused significant, although possibly short-term, shifts in the attitude and achievement of that student. Whether it was the treatment or a feeling of being special arising from being selected for the treatment cannot be known.

The second anecdote relates to the creative talents of another student from the experimental group. A writing assignment given in the regular classroom during the experimental period produced a complete, short text adventure, complete with alternate plots, maps and

diagrams, and a scheme for the reader to select their way through the story. This was done completely through the student's own interest.

Variables. The observations above suggest that spelling, vocabulary skills and student interest may play a significant part in the successful playing of text adventures. Since success equates to amount of content read, these variables may have an impact on the gains in reading comprehension through playing text adventure games.

Ten variables were measured which might have an impact on the gains in reading comprehension and reading sub-skills. Nine of these were listed in Table 1. They include age, sex, and average ability; achievement in vocabulary, reading, spelling, and language; difficulty of the text being read and attendance. The only other variable was homeroom class (and thereby teacher effect). In each measured variable the experimental and control groups were deemed to be equivalent. Differences between the experimental and control groups by any of these variables may therefore be attributable to the difference in treatment.

Age. Analysis by age was made by dividing the subjects into two groups, older and younger than the mean. The results of an analysis of gain scores for both experimental and control groups divided by age are presented in Table 5. Examination of the results shows that younger students performed significantly better than older students in structural analysis, inferential comprehension and total

comprehension. The bias towards younger students in inferential and total comprehension was also seen in the control group. It would thus appear that only the difference in structural analysis may be a result of the treatment.

Table 5
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Age Groups

Experimental Group					
	Older		Younger		
n =	9		13		
Reading Component	\bar{X}	σ	\bar{X}	σ	t-value
Phonetic Analysis	5	71	-48	107	1.298
Structural Analysis	8	27	43	53	-1.816**
Auditory Vocabulary	14	42	-29	109	1.126
Literal Comprehension	-14	31	-5	63	-0.381
Inferential Comp.	-32	56	23	71	-1.949**
Total Comprehension	-28	44	10	65	-1.516*
Control Group					
	Older		Younger		
n =	10		12		
Reading Component	\bar{X}	σ	\bar{X}	σ	t-value
Phonetic Analysis	-23	86	-6	46	-0.61
Structural Analysis	12	37	9	41	0.135
Auditory Vocabulary	11	59	-41	68	1.801**
Literal Comprehension	-10	74	9	52	-0.719
Inferential Comp.	-32	53	27	45	-2.796***
Total Comprehension	-24	56	21	30	-2.505***
*	p \leq 0.1		Key :		\bar{X} group mean
**	p \leq 0.05			σ	standard deviation
***	p \leq 0.01				
****	p \leq 0.005				

Sex. The results of a comparison of gains by sex are found in Table 6. For the experimental group there are no significant differences in gains for any of the sub-skills, or reading comprehension in total. This is in contrast to the control group, in which boys experienced significant gains over girls in auditory vocabulary and literal comprehension, while girls obtained superior gains in inferential

comprehension. There is insufficient evidence to conclude that playing text adventures is a less-sex-biased reading activity than USSR, but the results do play down the often held belief that computers are in some way a better learning device for boys than for girls (Eggers and Wedman, 1984).

Table 6
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Sex

Experimental Group		Male		Female		Unpaired t-value
n =		12		10		
Reading Component		X	σ	X	σ	
Phonetic Analysis		-15	128	-40	34	0.579
Structural Analysis		34	51	22	44	0.597
Auditory Vocabulary		8	79	-35	99	1.114
Literal Comprehension		-17	33	1	68	0.809
Inferential Comp.		11	86	-12	45	0.745
Total Comprehension		-5	62	-6	58	0.063

Control Group		Male		Female		Unpaired t-value
n =		14		9		
Reading Component		X	σ	X	σ	
Phonetic Analysis		-21	68	2	50	-0.9
Structural Analysis		14	40	4	39	0.62
Auditory Vocabulary		-3	65	-55	67	1.853**
Literal Comprehension		17	68	-20	35	1.491*
Inferential Comp.		-6	54	25	53	-1.36*
Total Comprehension		5	54	6	29	-0.067

*	p ≤ 0.1	Key :	X	group mean
**	p ≤ 0.05		σ	standard deviation
***	p ≤ 0.01			
****	p ≤ 0.005			

Ability. The ability variable was one of the hypotheses of the study, and the results of the analysis are given in Table 3. Low ability students gained more than high ability students in auditory vocabulary and literal comprehension,

while high ability students gained more than low in structural analysis and inferential comprehension. Comparison with the control group and the lack of overall significant F ratios meant that the null hypothesis for ability was accepted, however.

Vocabulary achievement. The four achievement variables were examined by dividing the experimental and control groups into above average and below average performance based on the mean of each achievement area. The results for vocabulary achievement are given in Table 7.

Table 7
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Vocabulary Achievement Groups

Experimental Group		Above Average		Below Average		Unpaired t-value
n =		10		12		
Reading Component		\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis		-34	125	-20	69	-0.341
Structural Analysis		44	60	15	29	1.476*
Auditory Vocabulary		-11	84	-12	97	0.028
Literal Comprehension		-7	45	-11	58	0.182
Inferential Comp.		24	92	-19	39	1.473*
Total Comprehension		9	70	-17	49	1.034

Control Group		Above Average		Below Average		Unpaired t-value
n =		13		10		
Reading Component		\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis		-10	75	-15	40	0.163
Structural Analysis		10	47	10	29	-0.047
Auditory Vocabulary		-47	70	8	57	-2.014**
Literal Comprehension		-6	64	13	54	-0.751
Inferential Comp.		16	44	-6	66	0.981
Total Comprehension		9	44	0	48	0.45

*	p <= 0.1	Key :	\bar{X}	group mean	
**	p <= 0.05		σ		standard deviation
***	p <= 0.01				
****	p <= 0.005				

In the experimental group, students having above average vocabulary skills had significantly greater gains in structural analysis and inferential comprehension than those with below average skills. This pattern was not mirrored in the control group, where students with above average vocabularies had significantly smaller gains than those whose skills were below average. A relationship between vocabulary and reading achievement gains was predicted on the basis of observations of the experimental group. Therefore these results may mean that vocabulary skills are a strong predictor of who will learn best through interactive fiction on computers.

Reading achievement. The results of analysis of the groups by reading achievement are presented in Table 8. It can be seen that students in the experimental group exhibiting above average reading skills had very significantly higher gains in inferential comprehension than those with below average reading skills. This difference was not found in the control group. Thus, the treatment may well have resulted in the gains in inferential comprehension for the above average readers of the experimental group.

Students with below average reading skills in both the experimental and control groups showed higher gains in auditory vocabulary than those with above average skills. As a result, the treatment was unlikely to account for the gains for below average readers.

Table 8
Comparison of Mean Gains in Reading Comprehension and Reading Sub-Skills by Reading Achievement Groups

Experimental Group

	Above Average		Below Average		
n =	12		10		
Reading Component	\bar{X}	σ	\bar{X}	σ	Unpaired t-value
Phonetic Analysis	-28	129	-24	42	-0.094
Structural Analysis	34	60	22	26	0.569
Auditory Vocabulary	-35	66	17	106	-1.415*
Literal Comprehension	-8	40	-10	65	0.109
Inferential Comp.	34	62	-40	57	2.897***
Total Comprehension	15	47	-30	66	1.86**

Control Group

	Above Average		Below Average		
n =	13		10		
Reading Component	\bar{X}	σ	\bar{X}	σ	Unpaired t-value
Phonetic Analysis	-17	71	-6	50	-0.409
Structural Analysis	8	46	12	30	-0.246
Auditory Vocabulary	-53	72	15	42	-2.647***
Literal Comprehension	-10	64	17	52	-1.089
Inferential Comp.	15	48	-5	63	0.845
Total Comprehension	7	46	3	46	0.201

* $p \leq 0.1$ **Key :** \bar{X} group mean
 ** $p \leq 0.05$ σ standard deviation
 *** $p \leq 0.01$
 **** $p \leq 0.005$

Spelling achievement. Spelling skills were observed to have an impact on the ability of students to be successful in playing adventure games, and thus were predicted to have an effect on reading achievement. Examination of Table 9 shows that students with above average spelling skills gained significantly more in inferential comprehension and total comprehension (explained by the gain in inferential) than students with below average spelling skills. However, the same gain in inferential comprehension occurred in the control group. As a result, no conclusion can be reached

with regard to spelling ability and reading comprehension growth in the experimental group from this study.

Table 9
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Spelling Groups

Experimental Group

	Above Average		Below Average		Unpaired t-value
n =	12		10		
Reading Component	\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis	-29	127	-23	43	-0.142
Structural Analysis	21	53	38	39	-0.854
Auditory Vocabulary	-24	73	4	107	-0.733
Literal Comprehension	-12	29	-5	72	-0.32
Inferential Comp.	32	62	-38	61	2.672***
Total Comprehension	11	48	-25	68	1.426*

Control Group

	Above Average		Below Average		Unpaired t-value
n =	12		11		
Reading Component	\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis	-11	52	-14	72	0.107
Structural Analysis	17	36	2	43	0.897
Auditory Vocabulary	-30	74	7	51	-2.138**
Literal Comprehension	-6	53	10	67	-0.631
Inferential Comp.	22	43	-11	62	1.502*
Total Comprehension	10	37	0	53	0.56

<p>* $p \leq 0.1$</p> <p>** $p \leq 0.05$</p> <p>*** $p \leq 0.01$</p> <p>**** $p \leq 0.005$</p>	<p>Key :</p> <p>\bar{X} group mean</p> <p>σ standard deviation</p>
--	--

Language achievement. The results with respect to overall language skills and gains in reading comprehension are presented in Table 10. Language achievement combines vocabulary, reading and spelling, and therefore the results in Table 10 should reflect those in the previous three. The

results show that students with above average language skills in both the experimental and control groups gained

Table 10
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Language Skills Groups

Experimental Group n =	Above Average 12		Below Average 10		
Reading Component	\bar{X}	σ	\bar{X}	σ	Unpaired t-value
Phonetic Analysis	-22	125	-32	48	0.249
Structural Analysis	31	60	25	26	0.315
Auditory Vocabulary	-42	95	25	69	-1.864**
Literal Comprehension	-12	58	-5	45	-0.32
Inferential Comprehension	21	73	-24	59	1.574*
Total Comprehension	4	67	-17	49	0.845

Control Group n =	Above Average 13		Below Average 10		
Reading Component	\bar{X}	σ	\bar{X}	σ	Unpaired t-value
Phonetic Analysis	-29	66	9	49	-1.502*
Structural Analysis	15	44	4	33	0.642
Auditory Vocabulary	-61	63	26	38	-3.852***
Literal Comprehension	-7	63	14	56	-0.83
Inferential Comp.	24	45	-17	59	1.885**
Total Comprehension	12	43	-3	49	0.797

*	$p \leq 0.1$	Key :	\bar{X}	group mean
**	$p \leq 0.05$		σ	standard deviation
***	$p \leq 0.01$			
****	$p \leq 0.005$			

significantly more in inferential comprehension, and that the students with below average language skills gained significantly more in auditory vocabulary than their counterparts. Since this result was consistent in both the experimental and control groups, it appears that overall language skills were not a predictor of gains in reading comprehension through playing computer based interactive fiction.

Text difficulty. A key component of USSR programs is the student's right to select his or her own text to read. This was permitted for both the experimental and control groups in this experiment, although the experimental group were limited to the four available text adventures. As a result, the difficulty level of the text being read was not consistent either within or between the experimental and control groups. An analysis of reading comprehension gains by text difficulty is given in Table 11.

Table 11
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Text Difficulty Groups

Experimental Group						
n =		Higher 16		Lower 6		Unpaired t-value
Reading Component		\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis		-23	51	-35	176	0.25
Structural Analysis		20	28	51	77	-1.416*
Auditory Vocabulary		-22	95	16	70	-0.875
Literal Comprehension		-13	51	1	57	-0.534
Inferential Comp.		16	67	-42	63	1.834**
Total Comprehension		2	59	-24	62	0.929
Control Group						
n =		Higher 7		Lower 14		Unpaired t-value
Reading Component		\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis		-30	77	-2	57	-0.922
Structural Analysis		7	43	10	42	-0.136
Auditory Vocabulary		-23	80	-23	66	0.007
Literal Comprehension		-23	66	20	56	-1.571*
Inferential Comp.		-4	74	18	43	-0.876
Total Comprehension		-13	53	21	35	-1.766*
* ** *** ****	p <= 0.1 p <= 0.05 p <= 0.01 p <= 0.005	Key :		\bar{X} σ	group mean standard deviation	

The difficulty level of the texts between the experimental and control groups was seriously skewed. Most students in the experimental group read text of high difficulty, while most in the control group read books of low difficulty. Therefore it is possible that text difficulty may have biased the overall results.

An examination of Table 11 reveals that students reading high difficulty text in the experimental group gained significantly more in inferential comprehension than those working with lower difficulty texts, while low text difficulty readers gained significantly more in structural analysis. These results were not found in the control group, where low difficulty readers gained significantly more than high difficulty readers in literal comprehension and total comprehension (explained by literal comprehension).

These results are difficult to interpret. It may be that higher difficulty text interactive fiction leads to the development of inferential comprehension. It may also be that those students who were involved with the higher reading difficulty text adventure had other characteristics which might explain the gain in inferential comprehension.

Table 12 gives the correlation co-efficients between text difficulty and the other variables. It can be seen that neither ability nor achievement aligns with text difficulty to suggest an alternative explanation for the gains in inferential comprehension. It appears, therefore, that there was a significant positive relationship between the reading difficulty

level of text adventures and growth in inferential comprehension.

Table 12
Correlation Co-efficients between
Text Difficulty and Selected Variables

Variable	Correlation
Age	-.169
Vocabulary	-.052
Reading	-.092
Spelling	-.019
Language	.019
Verbal Ability	.177
Ability Average	.004

Attendance. The attendance rate of all students involved in the study was quite high. Results of an analysis of gains in reading comprehension by high and low attendance groups are given in Table 13. Examination of the results shows that there were no significant differences in gains within the experimental group. In the control group, a significant difference existed in gains in inferential comprehension between those having high attendance and the average attendance group ($t_{hi:avg} = 2.32$, $p \leq .025$). However, the actual difference in attendance between the two groups was two to three classes. This minor difference in time cannot be a cause of the difference found.

On the basis of these findings, it appears that there was no difference in performance due to attendance.

Table 13
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Attendance Groups

Experimental Group					
Reading	n =	High	Average	Low	
Component					F ratio
Phonetic Analysis		-38	-30	6	0.269
Structural Analysis		26	38	8	0.598
Auditory Vocabulary		15	-15	-57	0.886
Literal Comprehension		-6	-6	-22	0.149
Inferential Comp.		-35	12	42	2.085
Total Comprehension		-22	0	16	0.59
Control Group					
Reading	n =	High	Average	Low	
Component					F ratio
Phonetic Analysis		-22	-17	39	1.237
Structural Analysis		16	15	26	1.563
Auditory Vocabulary		-5	-22	-88	1.409
Literal Comprehension		8	9	-42	0.93
Inferential Comp.		-22	32	20	2.988*
Total Comprehension		-6	20	-5	0.882
* $p \leq 0.1$					
** $p \leq 0.05$					
*** $p \leq 0.01$					
**** $p \leq 0.005$					

Homeroom class. Differences due to different homeroom classes for the subjects were experimentally controlled through randomization. Despite this, differences in daily instruction could result in differences within the experimental and control groups, so this variable was included for analysis. Results for gains in reading comprehension by homeroom class are contained in Table 14.

The actual treatment was given to the groups as shown in the table. Any difference in treatment application would therefore appear here.

Table 14
Comparison of Mean Gains in Reading Comprehension
and Reading Sub-Skills by Homeroom Class

Experimental Group		Class 5A		Class 5B		Unpaired t-value
n =		11		11		
Reading Component		\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis		-24	129	-28	52	0.089
Structural Analysis		32	61	25	29	0.362
Auditory Vocabulary		-37	99	14	74	-1.368*
Literal Comprehension		-12	62	-6	42	-0.246
Inferential Comp.		7	70	-7	72	-0.456
Total Comprehension		-1	70	-10	50	0.32

Control Group		Class 5A		Class 5B		Unpaired t-value
n =		12		11		
Reading Component		\bar{X}	σ	\bar{X}	σ	
Phonetic Analysis		-11	76	-14	44	0.1
Structural Analysis		8	47	12	31	-0.234
Auditory Vocabulary		-33	79	-13	57	-0.695
Literal Comprehension		-8	64	13	55	-0.823
Inferential Comp.		3	54	11	57	-0.347
Total Comprehension		0	40	11	51	-0.572

*	p <= 0.1	Key :	\bar{X}	group mean
**	p <= 0.05		σ	standard deviation
***	p <= 0.01			
****	p <= 0.005			

The results in Table 14 indicate that in auditory vocabulary, students in the experimental group from class 5B had significantly higher gains than those from 5A. This may be explained by the overall trend of students with lower than average achievement outperforming the higher achievement groups in auditory vocabulary seen in the previous tables, and the fact that a disproportionate number of the below average achievement students were in class 5B (t-tests for vocabulary, reading and language achievement revealed significant differences in favour of 5A).

Other than auditory vocabulary, the gains between members of the two halves of the experimental and control groups were the same. It would therefore appear that there was no real difference in the results on the basis of class of origin, and apparently no measurable difference in the application of the experimental and control treatments.

Table 15
Correlation Co-efficients for
Selected Variables by Reading Components

Reading Component	Group	Sex	Age	Ability	Vocab	Achievement Read.	Spell.	Lang.	Text Diff
Phonetic	E	-.128	.201	-.108	-.153	-.113	-.088	.064	.056
	C	.193	.028	-.218	-.059	-.222	-.087	-.319	-.406
Structural	E	-.132	-.311	.379	.314	.304	.067	-.016	-.032
	C	-.134	.092	.045	.163	.03	.186	.148	-.269
Auditory	E	-.242	.217	-.366	-.236	-.413	-.408	-.543	-.192
	C	-.375	.187	-.441	-.485	-.634	-.592	-.537	-.034
Literal	E	-.178	-.003	-.217	-.247	-.229	-.226	-.142	-.119
	C	-.309	-.302	.017	-.224	-.432	-.218	-.184	-.394
Inferential	E	-.164	-.536	.458	.307	.461	.365	.479	.379
	C	.285	.416	.43	.162	.073	.381	.383	-.389
Total Comp	E	-.014	-.351	.209	.069	.21	.118	.259	.203
	C	.015	-.502	.356	.022	-.163	.138	.166	-.50

Key : E = experimental group
C = control group

Summary of comparisons by variables. The variation in the effects of the variables on reading components is summarized in Table 15, which gives comparison correlation co-efficients for the experimental and control groups, by variable, for each reading component. The strongest

divergences are found in structural analysis and inferential comprehension. The strongest correlations are found in auditory vocabulary, and they show gains for the students with lowest ability and lowest achievement in both the experimental and control groups.

Figure 2 is an attempt to illustrate the findings described in detail in the sections above. It shows which half of each of the divided experimental and control groups experienced higher gain scores, by variable and reading component.

Figure 2
Comparison of Reading Comprehension and Reading Sub-Skills
by Variables Yielding Statistically Higher Gain Scores

	Group	Sex	Age	Ability	Vocab	Read.	Spell.	Lang.	Text Diff	Attend
Phonetic	E C									
Structural	E C	L	H	H						L
Auditory	E C	M	H	L H	L L	L L	L L	L L		
Literal	E C	M		L						L
Inferential	E C	F	L L	H	H	H	H H	H H	H	
Total Comp	E C		L L			H	H			L
Key :	E	experimental group			M	males		H	higher than average	
	C	control group			F	females		L	lower than average	

The chart shows that three components of reading comprehension (structural analysis, auditory vocabulary, and inferential analysis) were most strongly affected. Of these,

only the gains in structural analysis can be said to be unique to the experimental treatment group. On this basis, the findings related to the four hypotheses of this study are upheld.

Figure 2 also suggests that computer based interactive fiction may have provided a stronger vehicle for growth in inferential comprehension than USSR for students with above average ability and achievement. To explore this possibility, t-tests were conducted comparing the gains in inferential comprehension between the above average students of experimental and control groups by selected variables. A similar comparison was conducted between the below average groups separately. The results appear in Table 16 below.

Table 16
T-test Results Comparing Mean Scores of Experimental and Control Groups in Inferential Comprehension by Selected Variables

	Above Average Groups	Below Average Groups
Average Ability	0.438	0.011
Vocabulary Achievement	0.257	-0.568
Reading Achievement	0.866	-1.049
Spelling Achievement	0.453	-1.007
Language Achievement	-0.137	-0.294
Text Difficulty	0.644	-2.19**

- * p ≤ 0.1
- ** p ≤ 0.05
- *** p ≤ 0.01
- **** p ≤ 0.005

The t-test results in Table 16 show that overall, there was no significant difference in gains in inferential comprehension for either above average or below average

students regardless of treatment. The only significant difference was found for students reading below average text difficulty materials, in which the control group gained more in inferential comprehension than the experimental group.

Thus, computer assisted interactive fiction may be an effective way of developing inferential comprehension among above average ability and achievement students, but it cannot be claimed to be more effective than USSR.

Summary

The findings of the hypotheses can be summarized as follows:

- 1) No significant difference was found in gains in overall reading comprehension between students participating in computer based interactive fiction and those engaged in uninterrupted sustained silent reading;
- 2) Involvement in computer based interactive fiction was found to result in larger gains in the structural analysis sub-skill than participation in uninterrupted sustained silent reading;
- 3) No significant difference was found in the gains in overall reading comprehension or reading sub-skills between students of different average ability through the use of computer based interactive fiction.

The major finding of this study was thus that there seemed to be no difference in the use of computer based interactive

fiction or books in terms of the development of reading comprehension over the experimental period.

In addition to the findings of the hypotheses, it was found that computer based interactive fiction had a significant effect on the development of inferential comprehension. Students involved in the experimental treatment with above average ability and achievement levels seemed to gain in inferential comprehension more than those who were below average. There also seemed to be a positive relationship between the reading difficulty of the interactive fiction and the development of inferential comprehension.

The value of the data itself was questioned at the outset, due to the very high standard deviations associated with all mean gain scores. The findings, the quality of the data that they are based on, the conclusions that can be drawn from them, and potential new research arising from this study will be discussed in Chapter 5.

Endnotes

1. The parser is the mechanism in a text adventure game that examines what the player types into the computer in order to find key words that the game understands as commands, which causes the advancement of the plot.

Chapter 5 -- Summary and Conclusions

Summary

The purpose of this study was to examine the impact of student participation in computer text adventure games on the reading comprehension of grade five students over an eight week period. The study arose in response to the small number of studies dealing with the impact of computer based learning on achievement and the associated unclear picture that exists based on research, the desire to document new forms of computer based learning which are closer to the whole language approach to reading instruction than existing CAI programs, and the need to provide a research base for claims related to the value of computer based interactive fiction as an aid in the development of reading skills.

A review of the literature determined that the significant factors that affect reading comprehension in students included the availability of appropriate schemata, knowledge of story grammars and interest in the material being read. Computer assisted instruction in reading was seen to involve a host of sub-skills that do not touch on these characteristics to a large degree. In addition, existing CAI programs were found to reflect the reductionist model of language learning. In response to this many proponents of the whole language model of language learning have proposed the use of computer based interactive fiction as an

alternative to computer based drill and practice exercises that may accomplish the same goals, but do so in a manner more consistent with the whole language view.

Main Features of the Method

The experiment followed a true experimental pre-test post-test comparison group design. Sample and test instrument selections were made in order to conduct the experiment under real classroom conditions and make it as easily replicable as possible.

The sample of this study was composed of forty-five grade 5 students, from two classes in one elementary school. The subjects were divided into experimental and control groups by random assignment. Recent achievement test results were available for the sample from the local school district. All subjects were pre-tested with the CCAT Ability test (Verbal and Non-Verbal batteries), and Form A of the Stanford Diagnostic Reading Test (Brown Level).

The experimental treatment involved the experimental group of students in playing text adventure games on Apple //e microcomputers for 15 minutes per day, each day for an eight week period. This was an individual activity in which students worked essentially alone, although they could request assistance when necessary. Assistance took the form of asking questions to help the student make appropriate decisions. No other form of guidance was provided. A comparison treatment was administered to the control group

in the form of uninterrupted sustained silent reading.

Following the eight-week period, post-testing was conducted by administering the Stanford Diagnostic Reading Test (Brown Level) Form B.

Main Findings of the Study

The findings of the hypotheses are summarized in Table 17 below:

Table 17
Summary of Hypotheses and Decisions

	Hypothesis	Decision	Basis
1	Students engaged in computer based interactive fiction will not achieve significantly different gain scores in overall reading comprehension than students who engaged in uninterrupted sustained silent reading for equivalent time periods.	Accept	$t = -0.684$ not significant
2	Students engaged in computer based interactive fiction will not achieve significantly different gain scores in any reading comprehension sub-skill than students who engaged in uninterrupted sustained silent reading for equivalent time periods.	Reject	Structural Analysis $t = 1.431$ ($p \leq .1$)
3	Students of different ability levels engaged in computer based interactive fiction will not vary significantly in overall reading comprehension gains as measured by pre-and post-test reading comprehension scores.	Accept	$F = 1.128$ not significant
4	Students of different ability levels engaged in computer based interactive fiction will not vary significantly in gains in any reading comprehension sub-skill as measured by pre-and post-test reading comprehension sub-skill scores.	Accept	F ratios not significant

In addition, it was found that students using computer based interactive fiction of higher reading difficulty had higher gains in inferential comprehension than students working

with lower difficulty texts. Overall, students of above average ability and achievement had significantly higher gains in inferential comprehension than those of lower ability and achievement, but not more than comparable students engaging in USSR.

Based on the above, the most significant finding was that there appeared to be no significant overall difference between computer based interactive fiction and uninterrupted sustained silent reading in the development of reading comprehension.

Conclusions Based on the Findings

Before any discussion of the validity or generalizability of the findings above can occur, it is necessary to examine the nature of the data on which the findings were based, in light of the concerns raised in the analysis of Chapter 4.

Quality of the data. Examination of the data in all tables reveals huge standard deviations with respect to the size of the means for the gains results. This is underscored in the figures accompanying Figure 1. Typically, scores range nearly 200 points around means approximating zero. Such huge variances cast strong doubt on the validity of any findings or conclusions arising from this study.

A comparison of the standard deviations of the original testing data to their respective means shows that for each set of tests, the standard deviation was rarely in excess of 0.20 times the size of the mean. However, the standard deviation

of the gain scores ranged from just over 1 to 193 times the size of the mean, as shown in Table 18. This underscores the problem of comparing the mean value of the gain scores.

Table 18
A Comparison of the Size of
Standard Deviations to Means of Gain Scores

Variable	Experimental			Control		
	\bar{X}	σ	σ/\bar{X}	\bar{X}	σ	σ/\bar{X}
Phonetic	-26.27	96.24	-3.66	-12.17	61.24	-5.03
Structural	28.32	46.75	1.65	9.96	39.16	3.93
Vocabulary	-11.50	88.91	-7.73	-23.09	68.83	-2.98
Literal	-8.95	51.70	-5.78	2.09	59.42	28.43
Inferential	0.36	69.71	193.64	6.52	54.51	8.36
Comprehension	-5.41	59.33	-10.97	5.30	45.04	8.50

It would appear that the problem with gain scores is related to the comparability of the results obtained from Forms A and B of the Stanford Diagnostic Reading Test. This problem may have resulted because of the short duration of the experimental period, disparate conditions surrounding the giving of the two forms of the test, errors in scoring the test, or because of an inherent problem in the comparability of Forms A and B themselves.

The eight-week period of the experiment was short, possibly too short to produce gains large enough to be measurable by the treatments used in the study. However, one would expect in that event that the results of the pre- and post-tests would be very similar. In this case, they are not similar at all. While the net gains approximate zero, the actual variations in scores on a per-subject basis are very

large (the actual data is contained in Appendix A). This variation cannot be explained by the short duration of the experimental period.

One of the assumptions listed in Chapter 1 was that measurable reading comprehension gains would occur over an eight week period. This assumption may have been incorrect. The pre- to post-test results may have been as much a measure of the test-to-test stability of individual students as it was of gains in reading comprehension. Without treatmentless comparative data, this cannot be determined.

A second potential source of variability was the conditions surrounding the testing. The pre- and post-tests of reading comprehension were each given on the same day, at the same time, under identical conditions in classrooms which were adjacent. No unexpected events disrupted the giving of either test in any way. If problems existed with one of the test supervisors, it would show up in the results given in Table 14. T-tests carried out on the means and standard deviations of the two classes shown in Table 14 reveal that the means are not significantly different, but the standard deviations are: The mean standard deviation for 5A is greater than that for 5B ($t = 2.48$, $p \leq .025$).

This difference may be a result of a wider range of ability or achievement among the students of 5A. The average ability score for 5A was 105.57 with a standard deviation of 12; that for 5B was 95.41 with a standard

deviation of 15. This was found to be significant ($t = 2.42$, $p \leq .01$). Examination of the achievement averages yielded a similar pattern: T-tests on the achievement scores found 5A to have significantly higher means than 5B for all but spelling, where no difference was found. The students of 5A tended to be higher in ability and achievement than those of 5B, but not wider in range. It can therefore be concluded that there may have been a difference in testing conditions between the two classes that made up the sample of the study.

The third potential source of the problem lay in the scoring of the pre- and post-tests. However, the test scores were double checked when the gain scores were first calculated, and no errors in scoring were discovered.

The final potential source of the problem was the comparability of the two forms of the Stanford Diagnostic Reading Test. The SDRT examiner's manual indicated that the two forms were extensively checked for comparability prior to publishing, and alternate form reliability factors were given. The alternate form reliability was criticized by the Mental Measurements Yearbook (Mitchell, 1985) for being low (0.70 to 0.80), and for failure to provide specific test-retest information. An interesting factor was that the alternate form reliability was established with students from grades 7 and 8 only (Karlsen, Madden and Gardner, 1976). It may be that the reliability with lower grade levels was not as high as with the higher grade students.

On the basis of the above it appears that either the experimental timeline was too short to result in measurable outcomes, or some combination of test conditions and low alternate form reliability were the prime reasons for the wide range in the gain scores. Since the broad standard deviations existed in both classes of students (despite the fact that they were broader for 5A than 5B), it would appear that if the experimental timeline was not too short, the chief cause of the problem lay with the alternate forms comparability of the Stanford Diagnostic Reading Test.

Validity of the Findings. The extreme variances in the gains scores described above raise serious questions about the validity of the findings. The broad standard deviations mean that the score ranges of all comparison groups overlap to an enormous degree. In addition, the mean gain scores for virtually all tests reflect no growth, with approximately half of the students writing any given test showing regression rather than growth from the pre-test.

On the surface, it would appear that the findings of this study should have little validity. Certainly the findings must be replicated in a study free of the data problems found in this one prior to being accepted. However, the consistency of the range of scores across groups and tests, combined with the strength of the significance in the two cases that were found overall to be different from their comparison groups (the control group in the case of the second hypothesis, students reading high difficulty text adventures for the

finding regarding text difficulty) would suggest that these findings should not be wholly discarded.

The major finding of the study was that there was no significant difference in the development of reading comprehension of grade five students through the use of computer based interactive fiction or uninterrupted sustained silent reading. This finding would appear to be safe from the threats to validity discussed above, as the data for both groups were not only insignificantly different, but similar in pattern. It can thus be concluded that using computer based interactive fiction had the same effects on the readers as reading books would have. Teachers could therefore substitute playing text adventure games in place of reading books for the same duration as this experimental period, with some confidence that the students involved would receive similar benefits to reading books.

It was also seen that playing computer based interactive fiction seemed to develop the inferential comprehension of above average students more than below average students. Text adventure games may thus be an appropriate enrichment activity for above average students.

Generalizability. This experiment was designed to involve conditions as close as possible to actual classroom conditions. As a result, valid findings should be generalizable to the extent to which the sample subjects were representative of the population as a whole. In this case, the

students in the sample were found to be similar, and within normal bounds of ability and achievement.

The validity of findings of this study were questionable, due to problems in the comparability of the pre-test and post-test scores. As a result, no findings claiming significantly different performance in this study can be assumed to apply to any particular group other than the sample itself. They do, however, provide a basis for further research in the same vein.

The finding with regard to the similarity between computer based interactive fiction and USSR as a treatment is generalizable only within the eight week timeframe and short, daily application that was used in this study. It may be that longer term studies or studies using different lengths of time and regularity of treatment comparing text adventure game playing and USSR may find differences in comprehension effects. The extreme variability of the data prevents a prediction of what that effect might prove to be.

Potential Design Improvements

This study was undertaken to examine only the impact of playing text adventure games on the development of reading comprehension in the sample group. This may have been too narrow a focus for this kind of study.

Whole language proponents pointed to a range of benefits which playing text adventures might bring to language learning (Layton, 1987; Little, 1985; Jarchow and

Montgomery, 1985). One which was believed to be very significant was the interest which students show in playing these games, and its potential impact on reading skills.

The observations found that the interest of the students in the adventure games remained very high throughout the experimental period. The impact of interest on two special cases has already been described. The conclusion that can be drawn from this is that there may be something which is intrinsically motivating about text adventure games. This was the conclusion reached in the study undertaken by Lancy and Hayes (1986). Any future study of this kind should contain a measure of the student's interest in reading before and after the text adventure game experience, and a contrasting measure of their interest in the specific adventure game played as well.

A factor which may have been related to gains in reading comprehension through the experimental treatment may have been success in playing the adventure game itself. None of the students involved stopped reading, and none expressed a desire to quit the experimental group. However, there was a wide range in the actual success subjects experienced in completing the adventure game itself. Some students completed an adventure and started again (because it allowed for multiple plot patterns, in the case of MoonMist), some came near to completing, and some hardly got into the story by the end of the experiment. This last

group spent eight weeks basically re-reading the same text, with minor variations of sequence, over and over.

A success rating was not taken in this study due to the complexity of doing so. The many multiple plots make it difficult to figure out how successful a student has been, as one student may go immediately to a part of the plot that others may not get to until near the end of the story. One solution might be to keep a log book of where in the story each student starts, and where they end on a daily basis. This could be done through the use of the room indicator¹ on the monitor screen. The correlation between success rating and gains in reading comprehension might add considerably to the findings of a study such as this one.

The major problem in this study was the unreliability of the comparison of the pre- and post-test scores. To the extent that this arose from testing conditions, the problem could be rectified through careful training of the test administrators. The larger problem seems to have been the low reliability factor of the alternate forms themselves. A useful addition to this study would have been a third control group of similar students, who would not participate in any treatments, but who would write all tests, with the pre-test and post-test written without a significant time between writings. This would provide a rating of the comparability of the pre-test and post-test without treatment intervention. If time were available, it would be desirable to do this with a range of tests with students outside the experimental

environment prior to the final selection of the comprehension test instrument, in order to ensure that the instrument with the highest alternate form reliability was used in the study.

The use of gain scores as a means of determining growth over such a short experimental period may not be the most appropriate method available. A considerable amount of discussion has occurred with regard to the reliability of simple gain scores (Rogosa and Willett, 1983; Williams, Zimmerman and Mazzagatti, 1987). The concern is that the gain scores multiply the unreliability of the original pre- and post-tests, making their use questionable. Rogosa and Willett (1983) found that "when individual differences in true change are not small, the reliability of the difference score is respectable" (pp.339-340). In this study the gains were not only weak, but frequently negative and of extreme variation. As a result, the simple gain score may not have been the most appropriate means of determining the effectiveness of the treatments.

In sum, this experiment would have been greatly improved through the addition of measures of pre- and post-experiment student interest in adventure games, a measure of success in the adventure game played during the experiment itself, and pre-study validation of the comparability of the comprehension measure to be used (or at least simultaneous comparison group results). It might also have been helpful to have lengthened the timeframe of the study with a view to obtaining larger gain scores which

could have been compared with a higher degree of confidence.

Research Arising from the Study

A number of further research possibilities arise out of this study. First is the need to replicate the findings of the study itself, without the data problems that impinge on the validity and generalizability of its findings, in order to confirm the lack of difference between computer based interactive fiction and reading student selected books. A second, related study might include the measures of interest and success described above, to determine the relationship between those factors and gains in reading achievement, if any. In the event that a significant positive relationship were found between reading comprehension and success, it would be very important to ascertain which of the variables identified in this study may act as limiting factors of success. Vocabulary and spelling skills were postulated in this study, but not confirmed.

This study pointed to gains in structural analysis for students playing adventure games over those reading books, and to gains in inferential comprehension for those students engaged in interactive fiction of high reading difficulty over those using easier materials. These findings suffer most from the lack of validity in the data. A valuable study arising from these findings would be of the role of adventure games in the development of inferential comprehension, as on the

surface it would seem that the ability to infer is crucial to the successful playing of an adventure game. The theoretical link between these two is plausible. A study of the relationship might yield interesting results.

As defined in the endnotes to Chapter 3, structural analysis refers to the ability to separate words into their component roots and attachments. The link between playing adventure games and this skill is not obvious, and may have been a finding based on invalid data rather than a real occurrence. Further study would answer this question.

Once the answers to the questions above are obtained, there will be a need to replicate the study at different grade levels to determine if using adventure games in place of books is an acceptable alternative at particular grade levels, or if it is an optimum practice with students of a particular age.

If the lack of difference between the use of computer based interactive fiction and the reading of student selected books is confirmed, it would be interesting to know if there are particular students for whom adventure games are the most desirable form of reading. The study by Lancy and Hayes (1986) was an attempt to provide evidence for using adventure games to increase the reading interest in students who do not like to read. A direct study of this link is needed. This study found that in inferential comprehension, students who had high ability and achievement seemed to gain most from playing adventure games. It may be that adventure games are best suited to the brightest students.

Students with different learning styles may also gain differentially from different forms of reading (Pitts and Thompson, 1982). Again, further study is needed to see if any one kind of student would benefit most from playing adventure games.

It has been suggested that role-playing might develop schemata in students for unfamiliar or impossible contexts, such as dangerous, mythological or historical settings (Mavrogenes, 1983). Since adventure games are very complex role playing situations, they may provide an ideal vehicle for developing these kinds of schemata. Research is needed to confirm this.

Conclusion

Proponents of the whole language approach to language learning have suggested that the use of computer based interactive fiction in classrooms may be an appropriate alternative to computer assisted instruction drill-and-practice and tutorial exercises in developing student reading skills, in a fashion suitable to the whole language approach. This study has found that overall, the use of computer based interactive fiction has no different effects on reading comprehension than equivalent daily periods of uninterrupted sustained silent reading over a short (eight-week) period of time. It was also found that students playing adventure games experienced greater mean gains in structural analysis than students involved in USSR, and that students playing

adventure games at a higher reading difficulty level had greater gains in inferential comprehension than students using adventure games of a lower reading difficulty. In addition, computer based interactive fiction may be an appropriate means of developing inferential comprehension among students with above average achievement or ability..

It would appear that the suggestion of using adventure games to develop reading skills is acceptable. Unfortunately, the findings of this study must be seriously questioned due to problems with the data used in the analysis. There is sufficient evidence in this study only to recommend the undertaking of another study to replicate the results and confirm the findings. If those findings are confirmed, computer based interactive fiction has the potential to become a very useful reading tool in the hands of all teachers.

Endnotes

- 1 The place where you are in an adventure game is called a 'room' even if it is outdoors. The name of the room is always given in the top left hand corner of the screen.

References

- Anderson-Inman, L. (1987). The reading-writing connection: Classroom applications for the computer (part II). **The Computing Teacher**, 14 (6), 15 - 18.
- Anelli, C.M. (1978). Computer-assisted instruction and reading achievement of urban third and fourth graders (Doctoral dissertation, State University of New Jersey (New Brunswick, 1977)), **Dissertation Abstracts International**, 38, 6662A.
- Asher, S.R. (1980). Topic interest and children's reading comprehension. In Spiro, R.J., Bruce, B.C. & Brewer, W.F. (eds.) **Theoretical issues in reading comprehension**. Hillsdale, N.J.: Lawrence Erlbaum Associates. 525 - 534.
- Balajthy, E. (1987). What does research on computer-based instruction have to say to the reading teacher? **Reading Research and Instruction**, 27, 54 - 65.
- Bass, G., Ries, R., & Sharpe, W. (1986). Teaching basic skills through microcomputer assisted instruction. **Journal of Educational Computing Research**, 2, 207 - 219.
- Bear, G.G. (1984). Microcomputers and school effectiveness. **Educational Technology**, 24 (1), 11 - 15.
- Becker, H. (1984). Computers in schools today: Some basic considerations. **American Journal of Education**, 93, 22 - 39.
- Blair, T.R., Rupley, W.H., & Jones, M.P. (1986). Microcomputers: Another false prophet? **Reading Research and Instruction**, 26, 58 - 61.
- Bork, A. (1984). Computers in education today - And some possible futures. **Phi Delta Kappan**, 66, 239 - 243.
- Bradley, V.N. (1985). Computers and reading instruction. In Gambrell, L.B. & McLaughlin, E.M. (Eds.), **New directions in reading: Research and practice**. Maryland: 1985 Yearbook of the State of Maryland International Reading Association. ED 255 885
- Brady, P. (1986). Computer simulations and reading instruction. **The Computing Teacher**, 13 (2), 34 - 36.
- Bryg, V. (1985). The effect of computer assisted instruction upon reading achievement with selected fourth grade children. Dissertation Abstracts International, 45. In **ERIC Document Reproduction Service** ED 259 310

- Castner, T.R. (1982). Computer assisted instruction in reading and language arts. Unpublished manuscript, University of Georgia. **ERIC Document Reproduction Service** ED 237 967
- Clark, R.E. (1983). Reconsidering research on learning from media. **Review of Educational Research**, 53, 445 - 459.
- Clark, R.E., & Leonard, S. (1985). **Computer research confounding**. Paper presented at the annual meeting of the American Educational Research Association, March 31 - April 4, 1985. **ERIC Document Reproduction Service** ED 263 876
- Coomes, P.A. (1986). The effects of computer assisted instruction on the development of reading and language skills (Doctoral dissertation, North Texas State University, 1985). **Dissertation Abstracts International**, 46, 3302A.
- Cooperman, K.S. (1985). An experimental study to compare the effectiveness of a regular classroom reading program to a regular classroom reading program with a computer assisted instruction program in reading comprehension skills in grades two through four (Doctoral dissertation, The American University, 1985). **Dissertation Abstracts International**, 46, 1234A.
- Crafton, L. 1982. Comprehension before, during and after reading. **The Reading Teacher**, 35, 293 - 297.
- Deutsch, T.S. (1985). Interest in reading: A test of Kintsch's model (Doctoral dissertation, City University of New York, 1985). **Dissertation Abstracts International**, 46, 1235A
- Easterling, B.A. (1983). The effects of computer assisted instruction as a supplement to classroom instruction in reading comprehension and arithmetic (Doctoral dissertation, North Texas State University, 1982). **Dissertation Abstracts International**, 43, 2231A.
- Eggers, J.R., & Wedman, J.F. (1984). The growing obsolescence of computers in education. **Educational Technology**, 24(7), 27 - 29.
- Freebody, P., & Anderson, R.C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. **Reading Research Quarterly**, 18, 227 - 294.
- Geoffrion, L.D., & Geoffrion, O.P. (1984/85). Comprehension. **Computers in Reading and Language Arts**, 2, 18 - 24.
- Gerrell, H.R., & Mason, G.E. (1983). Computer-chunked and traditional text. **Reading World**, 22, 241 - 246.
- Good, T.L., & Brophy, J.E. (1977). **Educational psychology: A realistic approach**. Toronto: Holt, Rinehart and Winston.

- Grabe, M., & Grabe, C. (1985). The microcomputer and the language experience approach. **The Reading Teacher**, **38**, 508 - 511.
- Guzzetti, B.J. (1984). The reading process in content fields: A psycholinguistic investigation. **American Educational Research Journal**, **21**, 659 - 668.
- Holmes, B.C. & Allison, R.W. (1985). The effect of four modes of reading on children's comprehension. **Reading Research and Instruction**, **25**, 9 - 20.
- Icabone, D.G., and Hannaford, A.E. (1986). A comparison of two methods of teaching reading words to fourth graders: Microcomputer and tutor. **Educational Technology**, **26** (2), 36 - 39.
- Jameson, D., Suppes, P., & Wells, S. (1974). The effectiveness of alternative instructional media: A survey. **Review of Educational Research**, **44**, 1 - 68.
- Jarchow, E., & Montgomery, J. (1985). Dare to use adventure games in the language arts classroom. **English Journal**, **74**, 104 - 106.
- Johnson, S. (1982). Listening and reading: The recall of 7- to 9-year olds. **Journal of Educational Psychology**, **52**, 24 - 32.
- Juel, C., & Holmes, B. (1981). Oral and silent reading of sentences. **Reading Research Quarterly**, **4**, 545 - 568.
- Kinzer, C.K. (1986). A 5 part categorization for use of microcomputers in reading classrooms. **Journal of Reading**, **30**, 226 - 232.
- Kuechle, N. (1983). Reading, writing and programming. **Computers in Reading and Language Arts**, **1**, 7 - 10.
- Kulik, J.A., Bangert, R.L., & Williams, G.W. (1983). Effects of computer-based teaching on secondary school students. **Journal of Educational Psychology**, **75**, 19 - 26.
- Lancy, D.F. & Hayes, B.L. (1986). **Building an anthology of interactive fiction**. Paper presented at the annual meeting of the American Educational Research Association, April 16 - 20, 1986.
- Layton, K. (1987). Interactive text adventures. **Media and Methods**, (March/April), 14 - 16.
- Leinhardt, G., Zigmund, N., & Cooley, W.W. (1981). Reading instruction and its effects. **American Educational Research Journal**, **18**, 343 - 361.

- Litman, G.H. (1978). Relation between computer-assisted instruction and reading achievement among fourth, fifth, and sixth grade students (Doctoral dissertation, Northern Illinois University, 1977), **Dissertation Abstracts International**, 38, 2003A.
- Little, D. (1985). Adventure games in the curriculum: Problems and potential. **ERIC Document Reproduction Service** ED 265 491.
- Manning-Dowd, A. (1985). The effectiveness of SSR: A review of the research. **ERIC Document Reproduction Service** ED 276 970.
- Mavrogenes, N.A. (1983). Teaching implications of the schemata theory of comprehension. **Reading World**, 22, 295 - 305.
- Mazor, A. & Yussen, S.R. (1983). Children's ability to draw inferences from text. Unpublished manuscript, Wisconsin Center for Education Research, Madison, Wisconsin. **ERIC Document Reproduction Service** ED 239 239
- McCombs, B.L., Eschenbrenner, A.J. Jr., & O'Neil, H.F. (1973). An adaptive model for utilizing learner characteristics in computer based instructional systems. **Educational Technology**, 13(4), 47 - 51.
- McMillan, J.H. & Schumacher, S. (1984). **Research in education: A conceptual introduction**. Toronto: Little, Brown and Company.
- Miller, L. (1984). Computers and the Language Arts. Paper presented at the Colloquium on Canadian Research in Reading and Language Arts in Canada, Lethbridge, Alberta, June 7-9, 1984. **ERIC Document Reproduction Service** ED 260 408.
- Miller, S.D. & Smith, D.E.P. (1984). Differences in literal and inferential comprehension after reading orally and silently. Paper presented at the annual meeting of the International Reading Association. **ERIC Document Reproduction Service** ED 243 090
- Mitchell, J.V. Jr. (ed.). (1985). **Mental measurements yearbook (Ninth Edition)**. Lincoln, NB: University of Nebraska Press
- Newman, J.M. (1984). Online: Reading, writing and computers. **Language Arts**, 61, 758 - 763.
- Nicholson, T. (1986). Reading is not a guessing game: The great debate revisited. **Reading-Canada-Lecture**, 4, 239 - 246.
- Niemiec, R., Samson, G., Weinstein, T., & Walberg, H.J. (1987). The effects of computer based instruction in elementary schools: A quantitative synthesis. **Journal of Research on Computing in Education**, (Winter), 85 - 103.

- Oakes, J. & Schneider, M. (1984). Computers and schools: Another case of "...the more they stay the same"? **Educational Leadership**, 42, 73 - 79.
- Ortmann, L.N. (1984). The effectiveness of supplementary computer-assisted instruction in reading at the 4 - 6 grade level (Doctoral dissertation, University of the Pacific, 1983), **Dissertation Abstracts International**, 45, 140A.
- Packard, E.B. (1987). Interactive fiction for children: Boon or bane? **School Library Journal**, 34, 40 - 41.
- Parry, J.D., Thorkildsen, R.J., Biery, T.M., & Macfarlane, C.A. (1985 - 86). Computer based instruction: The transition from research findings to teaching strategies. **Educational Research Quarterly**, 10, 30 - 39.
- Pitts, M.C., & Thompson, B. (1982). The influence of children's cognitive style on reading comprehension. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, Texas. **ERIC Document Reproduction Service** ED 223 996.
- Porinchak, P. M. (1984). Computer-assisted instruction in secondary school reading: Interaction of cognitive and affective factors (Doctoral dissertation, Hofstra University, 1983), **Dissertation Abstracts International**, 45, 478A.
- Pumfrey, P.D. (1977). **Measuring reading abilities: Concepts, sources and applications**. Toronto: Hodder and Stoughton
- Ragsdale, R.G. (1985). Research needs for computers in education. In J. Wilton (Ed.). **Proceedings of the 1985 AEDS / ECOO conference**. Toronto: Educational Computing Organization of Ontario.
- Rogosa, D.R., & Willett, J.B. (1983). Demonstrating the reliability of the difference score in the measurement of change. **Journal of Educational Measurement**, 20, 335 - 343.
- Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In Spiro, R.J., Bruce, B.C., & Brewer, W.F. (eds.). **Theoretical issues in reading comprehension**. Hillsdale, N.J.: Lawrence Erlbaum Associates. 33 - 58.
- Samson, G.E., Niemiec, R., Weinstein, T., & Walberg, H.J. (1985). **Effects of computer-based instruction on secondary school achievement: A quantitative synthesis**. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April 1985.
- Schaudt, B.A. (1983). Another look at sustained silent reading. **The Reading Teacher**, 36, 934 - 936.

- Schiffman, S.S. (1986). Software infusion: Using computers to enhance instruction. **Educational Technology**, 26 (1), 7 - 11.
- Schmitt, M.C., & O'Brien, D.G. (1986). Story grammars: Some cautions about the translation of research into practice. **Reading Research and Instruction**, 26, 1 - 8.
- Smith, F. (1982). **Understanding reading (3rd edition)**. Toronto: Holt, Rinehart and Winston.
- Taylor, N.E., & Connor, U. (1982). Silent vs. oral reading: The rational instructional use of both processes. **The Reading Teacher**, 35, 440 - 443.
- Thompson, B.J. (1980). Computers in reading: A review of applications and implications. **Educational Technology**, 20 (8), 38 - 41.
- Thompson, M.E. (1983). Beyond the computer: Reading as a process of intellectual development. Paper presented at the annual meeting of the North Central Reading Association, Minneapolis, Minnesota, October 28-29, 1983. **ERIC Document Reproduction Service** ED 236 555.
- Thompson, R.A. (1984). Computer assisted reading instruction research. Paper presented at the annual meeting of the International Reading Association. **ERIC Document Reproduction Service** ED 243 091.
- Whitelock, R.F. (1987). **The effects of computer assisted instruction on student achievement at the junior high level: A review of research**. Unpublished manuscript, Brock University, College of Education, St. Catharines, Ontario.
- Williams, R.H., Zimmerman, D.W., & Mazzagatti, R.D. (1987). Large sample estimates of the reliability of simple, residualized and base-free gain scores. **Journal of Experimental Education**, 55, 116 - 118.
- Winkler, J.D., Shavelson, R.J., Stasz, C., Robyn, A.E., & Feibel, W. (1985). Pedagogically sound use of microcomputers in classroom instruction. **Journal of Educational Computing Research**, 1, 285 - 293.

Uncited Statistical References

The following texts were used in the data analysis as sources for and explanations of statistical procedures, although no reference to them appears in the body of the thesis:

- Bruning, J.L. & Kintz, B.L. (1977). **Computational handbook of statistics (2nd edition)**. Dallas: Scott, Foresman and Company.
- Hill, J.E. & Kerber, A. (1967). **Models, methods and analytical procedures in education research**. Detroit: Wayne State University Press.
- Johnson, R.R. (1976). **Elementary statistics (2nd edition)**. North Scituate, MA: Duxbury Press.
- Weinberg, S.L. & Goldberg, K.P. (1979). **Basic statistics for education and the behavioral sciences**. Boston: Houghton Mifflin Company.

Instruments Used in the Experiment

- Cochard, S. (1986). **MouseWrite**. Santee, CA: Roger Wagner Publishing Inc.
- Karlsen, B., Madden, R., & Gardner, E.F. (1976). **Stanford diagnostic reading test**. Harcourt, Brace and Jovanovich.
- King, E.M. (ed.) (1981). **Canadian tests of basic skills: Multilevel edition** (Levels 9 - 14, Forms 5 & 6). Toronto: Nelson Canada Limited.
- Wright, E.N. (ed.) (1974). **Canadian cognitive abilities test** (Form 1, Levels A - F). Toronto: Nelson Canada Limited.

Interactive Fiction Used in the Experiment

MoonMist,
Wishbringer,
Seastalker,
Planetfall, all copyright © Infocom Inc,
55 Wheeler Street,
Cambridge, Massachusetts, U.S.A.

Appendix A

The following pages contain the test scores obtained from all tests conducted with the subjects of the sample. These scores were processed using the Excel spreadsheet from Microsoft Corporation. The statistics contained within the study itself were generated by sorting the test scores using Excel, and copying the appropriate sorted data to the Statview statistical analysis package, where t-tests, ANOVA's and other statistics were created.

			CTBS Achievement Data					CCAT Ability Data	
Sex	Cls	Grp	Age (Apr 1)	Vocab	Reading	Spelling	Language Comp.	Verbal	Non Verb
M	A	E	10.83	5.20	5.80	5.70	5.90	5.70	97 115
M	A	E	11.67	6.10	5.60	6.00	6.20	5.70	101 96
M	A	E	10.58	6.10	7.10	6.20	6.80	6.80	119 119
M	A	E	10.33	7.30	7.70	7.00	6.90	7.10	130 107
M	A	E	10.42	7.60	7.20	7.00	6.30	6.90	123 117
F	A	E	10.17	7.60	6.90	7.60	7.40	7.00	142 116
F	A	E	10.67	5.20	5.50	5.20	6.00	5.80	107 96
F	A	E	10.08	6.10	5.30	6.00	6.20	5.50	114 98
F	A	E	11.58	3.80	4.80	4.90	4.80	4.80	86 97
F	A	E	10.58	5.80	5.90	5.50	5.80	5.90	107 105
F	A	E	11.50	3.80	4.60	4.00	4.80	4.60	76 77
M	B	E	11.67	6.10	5.30	5.20	4.50	5.30	89 97
M	B	E	10.83	5.20	4.10	5.70	5.40	5.10	92 81
M	B	E	11.75	4.10	4.00	5.00	4.20	4.20	79 77
M	B	E	11.08	5.50	4.80	3.80	4.90	5.10	106 93
M	B	E	10.50	3.30	4.40	2.50	3.00	3.90	103 97
M	B	E	10.25	7.00	6.40	7.60	6.60	6.80	123 105
M	B	E	10.50	5.40	5.60	5.00	5.30	5.50	136 96
F	B	E	10.67	5.40	5.60	6.80	6.20	5.80	103 94
F	B	E	10.75	4.30	4.30	3.80	5.00	4.30	78 59
F	B	E	10.25	5.80	6.20	7.00	5.50	5.70	109 83
F	B	E	11.00	5.40	5.70	7.00	5.80	5.20	80 79
M	A	C	12.50	4.80	5.20	3.80	4.80	4.90	87 83
M	A	C	10.25	4.70	4.00	4.00	5.30	5.20	99 112
M	A	C	10.75	6.40	5.90	6.20	6.40	6.30	109 106
M	A	C	11.50	3.00	3.90	3.80	4.00	4.00	72 91
M	A	C	10.42	6.40	5.60	4.90	5.70	6.00	113 110
M	A	C	10.58	5.90	6.30	6.80	6.70	6.40	113 119
M	A	C	11.00	5.90	6.60	4.70	6.00	6.30	112 103
F	A	C	10.83	5.80	5.00	4.10	4.60	5.30	97 110
F	A	C	10.83	4.80	5.00	6.20	6.30	5.30	101 104
F	A	C	10.58	6.50	6.70	5.80	6.10	6.30	116 111
F	A	C	10.50	6.20	6.30	6.60	6.20	6.00	111 102
F	A	C	10.75	6.70	7.40	7.00	6.90	6.70	107 123
M	B	C	10.67	7.30	6.30	7.30	6.10	6.50	109 116
M	B	C	11.42	4.70	4.80	5.80	5.30	4.90	84 70
M	B	C	10.92	5.10	6.30	7.30	5.90	6.20	105 120
M	B	C	11.83	3.60	4.60	4.00	4.00	4.20	65 66
M	B	C	11.08	6.50	6.00	5.50	5.60	5.90	106 102
M	B	C	10.25	6.10	5.80	3.80	5.00	5.50	105 110
M	B	C	11.00	5.20	5.30	5.80	6.00	5.50	99 113
F	B	C	10.83	2.50	3.10	4.00	4.80	3.70	92 82
F	B	C	11.92	4.50	6.90	3.20	4.60		85 79
F	B	C	10.33	5.60	5.00	5.20	6.40	5.50	105 97
F	B	C	10.75	5.80	6.20	8.20	6.90	6.40	106 123

Pre-test: Stanford Diagnostic Reading Test Form A								
Average	Phonetic Analysis	Struct. Analysis	Auditory Vocab.	Literal	Reading Comp. Infer.	Reading Comp. Total	Reading Rate	Post-test Phonetic Analysis
106.0	601	552	484	505	501	500	470	779
98.5	617	552	573	505	510	504	501	611
119.0	601	524	685	692	614	632	540	699
118.5	617	629	633	758	730	764	657	625
120.0	775	518	615	692	730	710	644	425
129.0	775	647	599	758	672	710	584	699
101.5	663	568	532	758	730	764	573	625
106.0	637	568	599	459	538	492	528	625
91.5	530	477	522	513	538	522	501	489
106.0	735	568	573	522	571	541	551	739
76.5	493	383	493	451	501	472	515	462
93.0	485	524	615	459	538	492	415	539
86.5	554	509	562	432	476	450	551	532
78.0	507	463	466	442	476	455	528	489
99.5	579	518	573	540	559	546	415	553
100.0	538	453	562	451	476	460	486	425
114.0	570	544	685	616	598	596	573	611
116.0	538	501	493	562	519	536	470	511
98.5	663	577	502	540	559	546	595	599
68.5	466	383	456	352	423	357	501	472
96.0	663	501	552	576	584	575	595	561
79.5	522	493	446	467	459	460	435	481
85.0	376	448	512	505	501	500	515	327
105.5	589	537	532	432	476	450	415	561
107.5	589	568	585	540	598	563	501	625
81.5	440	428	552	370	437	381	501	489
111.5	589	560	522	550	538	541	551	611
116.0	735	629	599	540	548	541	321	667
107.5	695	560	585	644	636	632	573	504
103.5	546	530	502	562	571	563	515	625
102.5	617	537	484	505	484	492	486	611
113.5	589	647	573	644	598	606	627	539
106.5	538	647	562	616	538	563	540	625
115.0	637	568	585	758	672	710	627	625
112.5	637	613	685	692	614	632	606	643
77.0	362	489	512	459	484	468	617	518
112.5	617	537	552	562	548	551	627	569
65.5	392	468	423	406	437	405	470	439
104.0	538	505	599	513	548	526	595	611
107.5	562	524	512	489	548	513	595	525
106.0	735	613	522	522	501	508	486	667
87.0	530	472	502	467	389	432	453	532
82.0	570	530	475	505	538	517	617	569
101.0	589	501	532	489	519	500	540	546
114.5	735	560	532	644	571	588	584	699

: Stanford Diagnostic Reading Test Form B					Reading	Pre-Post Comparison - Gain		
Struct.	Auditory		Reading Comp.		Rate	Phonetic	Struct.	Auditory
Analysis	Vocab.	Literal	Infer.	Total		Analysis	Analysis	Vocab.
550	470	488	498	490	370	178	-2	-35
511	552	508	517	508	524	-6	-41	47
542	566	712	780	759	599	98	18	-126
635	678	712	780	759	756	8	6	-80
685	780	630	688	655	609	-350	167	88
655	632	778	730	759	618	-76	8	-126
635	516	606	628	613	637	-38	67	-242
580	516	478	552	508	524	-12	12	57
483	488	478	479	477	547	-41	6	-25
685	539	606	565	579	558	4	117	17
377	470	488	498	490	418	-31	-6	19
580	552	400	372	357	558	54	56	93
515	479	400	479	421	434	-22	6	47
479	420	468	428	445	511	-18	16	-22
550	539	518	528	518	497	-26	32	-1
495	597	488	489	486	386	-113	42	146
592	597	606	730	655	657	41	48	-19
559	516	518	565	536	547	-27	58	-46
550	452	562	528	542	609	-64	-27	-88
372	452	430	400	397	482	6	-11	100
550	516	540	608	566	558	-102	49	-60
495	470	439	468	445	569	-41	2	3
472	506	439	479	453	466	-49	24	1
503	497	488	414	453	450	-28	-34	65
580	497	588	578	579	647	36	12	-43
437	461	439	358	381	434	49	9	91
655	566	588	608	594	536	22	95	16
655	566	606	592	594	418	-68	26	26
499	539	518	565	536	609	-191	-61	-105
535	539	551	565	554	637	79	5	-23
605	470	498	540	513	547	-6	68	-35
605	566	606	608	603	668	-50	-42	-78
605	452	540	592	560	569	87	-42	-164
605	614	712	730	723	668	-12	37	-144
635	597	664	608	625	756	6	22	-95
503	506	420	414	397	511	-44	14	47
559	452	551	565	554	756	-48	22	-110
472	376	410	428	397	497	47	4	-30
580	566	508	540	518	647	73	75	53
491	516	606	608	603	609	-37	-33	27
635	539	630	578	594	524	-68	22	17
442	488	508	489	494	418	2	-30	21
535	516	478	468	472	569	-1	5	11
542	488	508	489	494	618	-43	41	-1
550	566	606	628	613	569	-36	-10	-78

cores					Reading	Attendance	
Reading Comp.			Reading		Difficulty	Days	Percent
Literal	Infer.	Total	Rate				
-17	-3	-10	-100		4	38	100
3	7	4	23		6	38	100
20	166	127	59		6	33	87
-46	50	-5	99		6	33	87
-62	-42	-55	-35		4	38	100
20	58	49	34		6	35	92
-152	-102	-151	64		6	37	97
19	14	16	-4		6	34	89
-35	-59	-45	46		6	29	76
84	-6	38	7		4	35	92
37	-3	18	-97		4	34	89
-59	-166	-135	143		4	38	100
-32	3	-29	-117		6	37	97
26	-48	-10	-17		6	38	100
-22	-31	-28	82		6	37	97
37	13	26	-100		6	35	92
-10	132	59	84		6	36	95
-44	46	0	77		6	36	95
22	-31	-4	14		4	38	100
78	-23	40	-19		6	38	100
-36	24	-9	-37		6	38	100
-28	9	-15	134		6	32	84
-66	-22	-47	-49		6	36	95
56	-62	3	35		3	38	100
48	-20	16	146		3	38	100
69	-79	0	-67		7	38	100
38	70	53	-15		3	37	97
66	44	53	97		2	38	100
-126	-71	-96	36		9	38	100
-11	-6	-9	122		3	28	74
-7	56	21	61		5	38	100
-38	10	-3	41		3	31	82
-76	54	-3	29		3	33	87
-46	58	13	41		4	37	97
-28	-6	-7	150			36	95
-39	-70	-71	-106			34	89
-11	17	3	129		2	38	100
4	-9	-8	27		2	35	92
-5	-8	-8	52		2	38	100
117	60	90	14		3	37	97
108	77	86	38		3	35	92
41	100	62	-35		4	37	97
-27	-70	-45	-48		4	38	100
19	-30	-6	78		2	38	100
-38	57	25	-15		3	33	87

Appendix B

The following pages contain a transcript of parts of the MoonMist and Wishbringer text adventure games which were used by the experimental group in the study.

Here begins a transcript of interaction with

WISHBRINGER

Interactive Fantasy for Beginners

Copyright (C)1985 Infocom, Inc. All rights reserved.

Wishbringer is a trademark of Infocom, Inc.

Release 68 / Serial Number 850501

Okay, what do you want to do now?

>look

Hilltop

You're on a hilltop overlooking the seaside village of Festeron.

To the south stands the Festeron Post Office. It's a little brick building with a neatly-trimmed lawn. The Post Office door stands invitingly open.

Roads run down the hill to the east and west. There's a signpost nearby.

Somebody inside the Post Office is calling you.

What next?

>go to the post office

Post Office

This is the lobby of the Festeron Post Office. The walls are lined with small, private mailboxes and wanted posters. A service counter runs along the entire length of the room.

Your boss, Mr. Crisp, is behind the service counter reading other people's postcards.

Mr. Crisp hides the postcards away as you enter. "Where have you been?" he barks angrily. "Daydreaming again, eh? I've been looking everywhere for you!"

(Your score just went up by 1 point! Your total score is 1 out of 100.)

What next?

>listen

At the moment, you hear nothing unusual.

Mr. Crisp reaches under the service counter and pulls out a mysterious envelope. "We just got this Special Delivery," he snarls, tossing it onto the service counter. "I want you to drop it off right away. That means NOW!"

What next?

>get the envelope

Taken.

(Your score just went up by 5 points! Your total score is 6 out of 100.)

Mr. Crisp drums his fingers on the service counter impatiently. "Deliver that envelope! The Magick Shoppe closes at five o'clock!"

What next?

>Crisp, where is the magick shoppe?

(To locate something, use the command: FIND THING.)

What next?

>find magick shoppe

How do you intend to do that?

Mr. Crisp growls, "Quit stalling!"

What next?

>ask mr. crisp where the magick shoppe is

(To locate something, use the command: FIND THING.)

What next?

>leave post office

As you leave the Post Office, somebody slams and locks the door behind you.

Hilltop

>s
The Post Office door is closed.

What next?

>e

Outside Cottage

You're outside the cottage of Miss Voss, the town librarian. The road turns north toward the village, and bends upward to the summit of Post Office Hill.

As you approach the cottage you are greeted by the tiniest poodle you've ever seen in your entire life. It charges across the street, yapping angrily and showing its sharp little teeth.

You can smell the poodle's hot breath as it yaps.

What next?

>w

The poodle yaps a threat as you back away.

Hilltop

What next?

>w

Outside Cemetery

You're standing next to an open iron gate that leads west into the Festeron Cemetery. A road runs east to the top of Post Office Hill.

<You won't see the "What next?" prompt any more.>

>w

The cemetery is a fearful place. Are you sure you want to go in there?

<Please type YES or NO.> >yes

You have been warned.

Creepy Corner

You're in a creepy corner of the Festeron Cemetery, surrounded by silent tombstones. An iron gate opens to the east, and a narrow lane wanders north.

>n

Spooky Copse

A copse of willow trees makes this part of the cemetery look really spooky. Narrow lanes wander south and west.

There's an open grave nearby, freshly dug, with a tombstone erected next to it.

An old gravedigger is resting under a willow tree.

The gravedigger nods a greeting as you approach.

>wait

Time passes.

The gravedigger notices the mysterious envelope you're holding. "That's a mighty mysterious-lookin' envelope you got there," he says. "Let's have a look at it."

>give envelope to gravedigger

The gravedigger peers at the address on the mysterious envelope. "Hmm," he mutters, handing it back to you. "Could've sworn I buried that old woman years ago."

<You are again holding the mysterious envelope.>

"Couldn't ask for a nicer day for diggin' graves."

>wait

Time passes.

"Got to go," says the gravedigger, picking up his shovel. "See you soon."

The old man ambles away to the west.

>look in grave

It's six feet deep and freshly dug. You can see an old bone inside.

>enter grave

Open Grave

You're at the bottom of an open grave, surrounded by six-foot walls of dirt.

There's an old bone here.

>get bone

Taken.

(Your score just went up by 1 point! Your total score is 7 out of 100.)

>leave grave

With great difficulty, you manage to climb out of the open grave.

Spooky Copse

>s

Creepy Corner

>e

Whew! That cemetery sure is spooky.

Outside Cemetery

>verbose

(Maximum verbosity.)

Outside Cemetery

You're standing next to an open iron gate that leads west into the Festeron Cemetery. A road runs east to the top of Post Office Hill.

>e

Hilltop

You're on a hilltop overlooking the seaside village of Festeron.

To the south stands the Festeron Post Office. It's a little brick building with a neatly-trimmed lawn. The Post Office door is closed.

Roads run down the hill to the east and west. There's a signpost nearby.

>e

Outside Cottage

You're outside the cottage of Miss Voss, the town librarian. The road turns north toward the village, and bends upward to the summit of Post Office Hill.

An angry poodle is blocking your path.

The poodle glares at you through eyes red with hatred.

>throw bone to poodle

The poodle tests the old bone with its tongue, lies down and begins to gnaw on it, keeping a red eye on you.

(Your score just went up by 3 points! Your total score is 10 out of 100.)

>n

The poodle yaps a threat as you back away.

Rotary South

This is the south side of the Festeron Rotary. A road branches south, towards Post Office Hill.

The Festeron Public Library, famous for its museum of local historic artifacts, stands proudly on the nearby corner.

Miss Voss, the town librarian, is locking the library door as you approach. "Just the person I was looking for!" she exclaims, smiling brightly.

>

Here begins a transcript of interaction with
Moonmist
Infocom interactive fiction - a mystery story
Copyright (c) 1986 by Infocom, Inc. All rights reserved.
Moonmist is a trademark of Infocom, Inc.
Release number 4 / Serial number 860918

What would you like to do?

>leave car

You are on your own feet again.

(You are now in the driveway.)

You are by the front gate of Tresyllian Castle. You can hear the ocean beating
urgently against the rocks far below.

In the moonlit gloom, you can make out an ornament on the gate. It's a winged,
two-legged dragon called a wyvern, which crests the Tresyllian family's coat of
arms.

The dragon appears in profile. The moonlight glints on its lone visible eye.

Your new little sports car is parked here.

What next?

>press the dragon's eye

The dragon's eye glows red. A voice comes from a hidden speaker. It says:

"Please announce yourself. State your title -- such as Lord or Lady, Sir or
Dame, Mr. or Ms. -- and your first and last name."

>Sir Loin d'Boeuf

"Did you say your name is Sir Loin D'Boeuf?"

>yes

"And what is your favorite color, Sir Loin?"

>blue

"Did you say your favorite color is blue?"

>yes

"Jolly good! The spare bedroom is decorated in blue! Please enter."

The red eye turns green, and the front gate creaks open.

What next?

>enter castle

The front gate closes and locks behind you.

(You are now in the courtyard.)

As flood lights blaze on, you look around. It looks even lovelier than it
sounds in the tourist brochure.

The dark stone turrets rise toward the misty sky.

Someone comes running out of the wing to greet you. She's a beautiful
red-haired young woman of average height. You recognize her as your friend,
Tamara Lynd.

"Loin!" she cries with outflung arms. "You sweet thing, to answer my letter in
person this way! And all the people I wrote about are here tonight for Lionel's
memorial birthday dinner!"

After a warm hug, she asks anxiously, "You did read my letter, and not just
give it a hasty glance?"

>yes

"Then you know about my engagement, and the White Lady, and the fact that...
that someone is trying to kill me!"

[You won't see "What next?" any more.]

>yes

She hugs you with affection. "I'm so glad you're here!"

>wait

[I assume you mean: Wait 10 minutes.]

Time passes...

"We can talk more later, Loin," says Tamara, taking your arm, "but let's go in
now, so you can meet the other guests. Bolitho will see to the car and bring
your luggage."

An elderly butler appears.

>examine butler